



UC DAVIS

**GRADUATE SCHOOL
OF MANAGEMENT**

Handbook on Writing Data-Driven Business Reports

FOR CURRENT AND PAST STUDENTS OF 203B AND 285

(PLEASE DO NOT CIRCULATE OR MAKE ANY COPY)

ISSUED: 6/10/2016

REVISED: 6/10/2017, 6/11/2018, 9/16/2018, 10/20/2020 AND 5/30/2021

PREPARED BY: RAYMOND BERISTAIN-BARAJAS

REVISED BY: ROBERT BOHN, TYLER BORUFF, JUSTIN CHEN, MICHAEL MIKHOV, BRANDON QUACKENBUSH, SACHEL SUMMERS, KYLE TARVIN, PRAYAS TIWARI & EUGENIA WILSON

PREPARED, SUPERVISED, INSTRUCTED AND REVISED BY: PROFESSOR CHIH-LING TSAI

Contents

| | |
|---|----|
| Preface | 3 |
| 1. Living with heart..... | 3 |
| 2. 3B's, CAMP, and CI | 3 |
| 3. SHE/HE/THEY | 3 |
| 1. Introduction | 4 |
| 2. Purpose and Motivation | 5 |
| 2.1. Internship & Job Search | 6 |
| 2.2. Converting a Summer Internship or Securing a Promotion..... | 6 |
| 3. Preparation | 6 |
| 3.1. Floor Plan | 6 |
| 3.2. Pitfalls..... | 8 |
| 4. Report Writing | 10 |
| 4.1. Structure | 10 |
| 4.2. Miscellaneous Tips | 18 |
| 5. Post-Report Writing | 18 |
| 5.1. Submission Etiquette | 18 |
| 5.2. Second Report / Report Modification..... | 18 |
| 6. Conclusion..... | 19 |
| Appendices..... | 20 |
| Appendix A: Data Resources..... | 20 |
| Appendix B: Example of Benchmark Forecasting | 20 |
| Appendix C: Report Template..... | 21 |
| Appendix D: Sample Business Report | 25 |
| References | 25 |
| 7. Recommendations for Instructing a Course Utilizing Data-Driven Business Reports..... | 26 |
| 7.1 Prior to First Day of Instruction..... | 26 |
| 7.2 On the First Day of Instruction..... | 27 |
| 7.3 Tracking Progress | 28 |
| 7.4 Final Project Preparation and Presentation..... | 30 |
| 7.5 Supplemental Materials..... | 31 |

| | |
|---|----|
| 8. Supplementary Material: Data Resources Handbook..... | 36 |
| 8.1 Individual Data Resources..... | 36 |
| 8.2 Data Repositories..... | 39 |
| 8.3 Data by Industry..... | 40 |
| 8.4 Book Resources..... | 42 |
| 8.5 Artificial Intelligence / Machine Learning Data Resources | 42 |

GSSM-UCD

Preface

As graduates (or soon to be graduates) of the UC Davis Graduate School of Management, we represent the school's community whether we intend to or not. Through outstanding work quality, we can elevate the perception of the UC Davis GSM MBA program, which ultimately benefits ourselves and future generation of UC Davis MBAs. Even if you don't remember what a univariate ARIMA Model or Winters' Method is, you should remember these few things from your time as a 203B and 285 student:

1. *Living with heart*

- Purify your heart
- Cool your heart
- Warm your heart
- Heart to heart

2. *3B's, CAMP, and CI*

- Live a **B**alanced life, **B**udget your time, and you will achieve your **B**est performance.
- **C**onfidence—**A**ppreciation—**M**orality—**P**ersistence
- **C**reative **I**ndependent Thinking

3. *SHE/HE/THEY*

- **S**mile, be **H**appy, and **E**njoy
- Stay **H**ealthy and **E**xercise

...most importantly, don't forget to explain statistics to grandma.

That's not too much to remember, right? As an MBA equipped with these three tenets, we can offer a valuable alternative mindset to the organizations we will work for throughout our careers. Furthermore, the ability to clearly communicate in writing will be an asset everywhere we work. Professor Tsai's teachings extend beyond the forecasting methods he covered in class. This Business Report Writing Handbook is a way for us to extend and retain our learning from 203B/285 beyond what we captured during our short time at the GSM. We hope this material is a positive contribution for all of you.

If you were a 285 student, this Handbook should serve as a reminder and a guide for future improvement. If you finished 203B but not 285, this Handbook can provide you some guidance on professional report writing. You may find the structure different from 203B. If you have questions, use your alumni network to help you. And don't forget to use the spirit of 203B!

Written communication skills are some of the highest demanded qualities for new MBA hires. A strong understanding of finance, marketing, and operations is crucial, but without the ability to communicate your work your impact can be severely limited. Good written communication skills require independent and critical thinking, and thorough understanding of your business. Remember, report writing is not the purpose but the tool that organizes and strengthens your insights. This Handbook will guide you on how to write a professional business report that can leave a lasting impression, improve your professional written skills, separate you from your competitors, and ultimately add value to your organization.

1. Introduction

The ability to write a succinct and professional business report is invaluable for students and graduates of MBA, Business MS and Business PhD programs. In practice, employers expect a high level of proficiency in creating and delivering data-driven verbal and written communications on business topics. Many business schools offer an introductory core course on business communication, comprised of presentations and writing. However, gaining the required skills to write a full-length business report takes solid discipline and real-life experience. This handbook will serve as a tool for MBA (or MSBA) students. For example, it could be used by students summarizing his/her summer internship work, or by students preparing for an interview. In either case, the report could be mandatory (e.g., fulfilling the manager’s assignment) or it could be a voluntary way of leaving a lasting impression on the potential employer.

A study by the National Association of Colleges and Employers, found that 70.2% of employers seek candidates who have strong written communication skills (NACE, 2016). Written communication skills ranked third as the most desired quality overall, behind leadership and the ability to work as a team member. A similar study, the GMAC Corporate Recruiters Survey (Figure 1), found that written communication skills was the third most demanded quality of new MBA hires, following oral communication and listening skills (GMAT, 2017). Even if you possess the skillset and knowledge of a business school graduate, without the necessary skills to communicate your work, it won’t add value to your company.

As an MBA student, you are accustomed to leadership workshops and a lot of teamwork, but how often are you required to write full business reports? Furthermore, another study found that businesses are spending \$3.1 billion annually on writing training, of which \$2.9 billion is spent on current employees (CollegeBoard, 2004). The ability to communicate clearly through writing is a skill that is highly valued in the work place. As Jason Fried, founder of Basecamp, put it, “Clear writing is a sign of clear thinking. Great writers know how to communicate. They make things easy to understand.” (Fried, 2010) As a great writer you will add value to your organization and your career will benefit.

What constitutes a business report? A critical component of most business reports is to analyze a trend or business problem and provide recommendations on how to alleviate the problem, or how to benefit

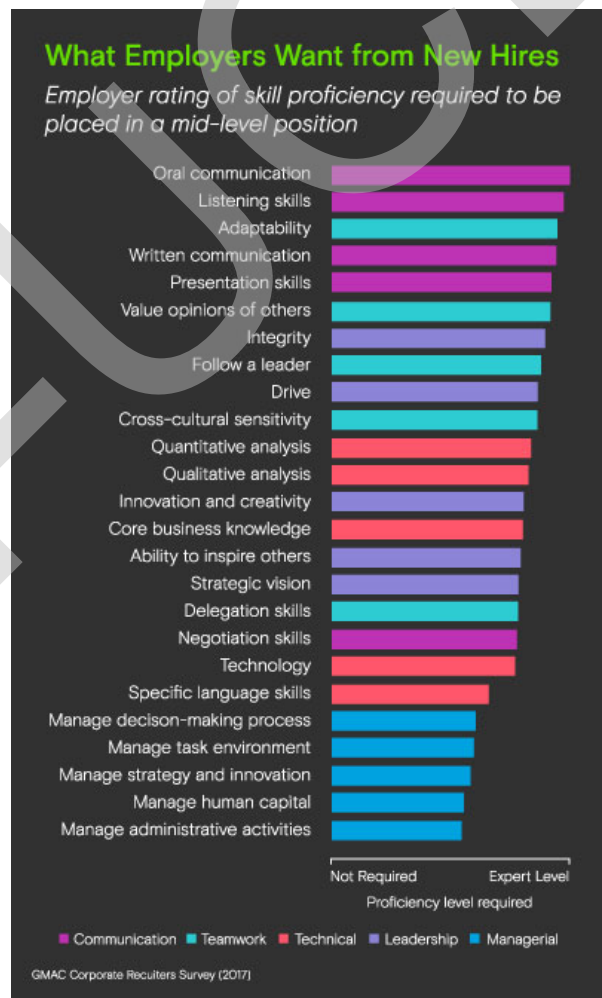


Figure 1: Demanded Skills for New MBA Hires

from a potential business opportunity. These reports are generally targeted at executives, or someone in a decision-making role, within the company.

Equipped with the tools in this handbook, you will be able to communicate all the skills you have learned throughout your MBA (or Business MS) education. Demand for business analytics is increasing. Experience analyzing “big data” will grow in demand with proportion to cloud computing and cloud storage. Data analysis and statistical modeling is fundamental, but having an analytical mindset extends beyond forecasting or financial modeling. Many aspects of a business require analytical skills, whether it is analyzing the effectiveness of teams within your organization as a human-resource professional, or forecasting the value of a potential new business venture as a financial analyst. The organization and logic behind business report writing extends to qualitative or empirical data. Teams in this category also benefit from a written business report.

We will discuss use cases of business report writing. You will be directed to websites for collecting large amounts of data on diverse topics in various industries. A comparison of data analytics software packages will be discussed. The general layout of a business report will be outlined. And finally, you will have access to a report template as a reference.

The goal of this handbook is to help MBA (or Business MS) students secure an internship or full-time position, and set them up for career success as business professionals. Writing an effective business report that makes the job of a decision maker easier is a differentiator that will not go unnoticed in a corporation. Writing a succinct report is a skill that requires guidance and practice. In addition to report writing skills, we hope to provide guidance to instructors that are interested in teaching data-driven reports in any classroom context.

The remainder of the report is organized into the following structure. Section 2 - Purpose and Motivation covers common situations when you need to write a report. Section 3 - Preparation talks about how to prepare for report writing and pitfalls that need to be paid attention to. Section 4 - Report Writing covers structure and writing skills. Section 5 - Post-Report Writing shows considerations after finishing the report. Section 6 – Conclusion illustrates, further, the motivation and necessity for students to utilize this handbook in the classroom and the workplace. Section 7 – Recommendations for Instructing a Course Utilizing Data-Driven Business Reports serves as a means to guide future instructors in facilitating a curriculum that seeks to integrate professional report writing. As well, it provides supplementary materials that instructors and Readers/TAs can utilize in the conduct of such a course.

2. Purpose and Motivation

There are many reasons why writing a business report is beneficial for *you*. Whether you are job hunting and want to impress the hiring manager, or currently employed and looking for a raise or promotion. Writing an effective business report will not only benefit the company, but it can differentiate you from others and help you achieve your career goals. As always, whenever negotiating a job offer or potential raise/promotion, the value you add to the organization should be at the forefront. A business report is one of the best ways of expressing the value that you add to the team. This is especially the case if writing business reports is not in your job description or is not particularly expected of someone in your position. As an MBA (or a Business MS) graduate, there will be many positions where you are expected to analyze a business problem or opportunity and present it to a group of people at your company.

2.1. Internship & Job Search

Business school is much like law school or medical school. These institutions prepare its students for careers in specific fields—in your case, business. Because of the nature of the MBA (or Business MS) program, there is a major emphasis on job search and career development. Many students have years of work experience and would like to change careers or switch industries. There are also others, who have limited work experience or may come directly from an undergraduate degree. In any case, the importance of job placement cannot be understated at a business school. Many business schools are measured by the graduating class' job placement rate three months after graduation, along with average salary/bonus/total compensation.

What many people don't fully realize is the extent that networking is critical to a business school education. As an MBA student, you are taught networking social norms to accompany your academic courses, but the emphasis is on social types of communication. There is a need for formal training in report writing communication for a business environment. Almost any MBA (or Business MS) career path benefits greatly from business report writing.

2.2. Converting a Summer Internship or Securing a Promotion

Perhaps you are already a full-time employee or summer intern at an established company. As an employee, when negotiating a promotion, remember that your job is to demonstrate the value you have added to the organization. An effective way to quantify this value is by writing a report on the work you have done. Having a tangible example of the quality of your work will go a long way. Similarly, as an intern looking to land a full-time offer, a written report can leave a lasting impression and tangible proof of your work once you return to school. Summer internships are often structured with a final presentation near the end to summarize your work, but this proof of work disappears once summer is over. A written report in conjunction with a final presentation can allow your work to be documented and easily transitioned to another employee if the company chooses.

By now you should be convinced of the importance of business reports and how they can benefit your career. You may have written a report and understand their impact, or you may not yet know where to start, but are interested to learn more. Below we will discuss how a business report can be structured, where you can access large sources of free data, which statistical software packages are at your disposal, and what kind of pitfalls you might have.

3. Preparation

Sometimes you may find that it's difficult to start a project. This section will give you some ideas on how to prepare for a report before you actually jump into it. It's important to have a floor plan before you start building a house.

3.1. Floor Plan

3.1.1. Purpose

As mentioned in the previous section, report preparers might have different reasons (interview, end of internship, promotion, etc.) to write a report. Knowing the purpose is important and can help you set the

tone. Be aware of the difference between this and the purpose of your report. Your purpose of writing a report might be conversion from an intern to a full-time employee. But your report's purpose should be the problem you want to address and solve in your report.

3.1.2. Audience

Who is your audience? You should keep this in mind when you're writing the reports. They could be your peers, interviewers, supervisors, and/or even executives in your company. Because each group has a different level of understanding on your topic, you need to present the information differently. Peers and supervisors might want to know more about details, while VPs and executives might be concerned about the big picture. HR, on the other hand, might want to see how you can fit in the company's culture. Knowing your audience can help you better organize the report.

3.1.3. Flow Chart

After identifying your purpose and audience, you need to build up a floor plan for your house. Figure 2 shows what we recommend you use to lay out your plan: backward thinking. You start with what you want to have at the end, a report in this case. Next, list out what you need to do to reach the end. You might need some sets of data, software, and other resources. And then you need to think about whether you have these already, where and how to gather this information, and whether you have alternatives. Finally, this flow chart should be consistent with your time table, which will be discussed in the next section.

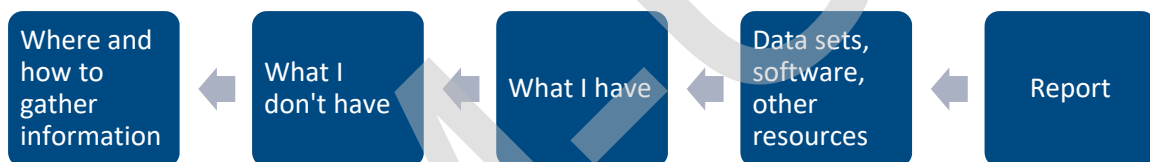


Figure 2: Flow Chart of Report

3.1.4. Time Table

Time is perhaps your biggest obstacle. Whether you are writing a report in preparation for an interview in three days, or you have decided to complete additional work by writing this report in your spare time, establishing a disciplined work flow can help you achieve an excellent report within your time constraints.

The importance of distraction-free work spaces cannot be understated. If you enjoy spending time with family, watching movies, or playing video games, you might not want to spend all day writing business reports. Instead of trying to accomplish many tasks at once, you would be surprised how much progress you can make by setting a time and place to write *exclusively*.

Use your time wisely. It might take a long time to find the useful data and make a good story. Occasionally, you may find yourself without good data or a good story. In that case, you need to decide what compromise you want to make. Relatively speaking, a good story with imperfect data is preferable, because readers usually want to see critical thinking from the writers, not searching skills. As long as the problems are not critical to your assumptions, you can discuss what problems the data has and what these problems might affect your recommendations. Lastly, give your first draft to someone who's not familiar with your topic and get some feedback, if time allows.

3.1.5. Data Sources

With the internet at your disposal, you have access to many repositories of large public datasets. See Appendix A for a list of helpful databases. Keep in mind who your audience is when using sensitive data. If certain data is confidential, it should remain confidential and only included in your report if the reader has authorization to read it. A report for an interview should always use public data to avoid any complications.

As an alternative, if you are capable and want to display another skillset that can separate you from your colleagues, an option is to collect your own data. Data collection can come from surveys, interviews, or focus groups/ethnographic studies that you conduct on your own. We only recommend taking this path if the job description requires data collection, otherwise using existing data and showcasing your analytical skills is more than enough. Remember, the purpose of writing a business report is to display your written communication skills, which includes your analytical mindset and business knowledge.

3.1.6. Software

The number of data analytics software packages available to you can be overwhelming. There are several factors that differentiate them: price, general-purpose v. specialized, and difficulty. Table 1 is a list of some software packages for business. It is important to do preliminary research on your company of interest, on which software package is preferred. Some job descriptions explicitly state their preference, while others use the generic terms “research methods” or “data analysis”.

Table 1 - List of Software Packages

| | | |
|----------------|---------------|-----------------------|
| Excel | Stata | Tableau |
| Minitab | Python | Raw |
| SPSS | SAS | Visualize Free |
| R | MATLAB | |

3.2. Pitfalls

This section will provide recommendations as to how to avoid pitfalls that are common in this type of writing. The two most common pitfalls that we will discuss below are: presentation of technical materials in the report and including output in figures or table format.

3.2.1. Technical Materials

One key aspect of report writing to consider is how much and how to present technical concepts (i.e., statistical terminology) in the report. One pitfall is that the writer assumes the reader is familiar with statistical concepts and includes a lot of technical elements. When deciding if you should add terminology, a good rule of thumb is to ask yourself, “Is it vital?” If you can make your point without presenting terminology, then you should do so. If you do need to present a technical discussion, a good approach to follow is to include a reference of the terminology in the footer. This way the reader will have an understanding of what you’re talking about and can research the topic further should he/she find it necessary.

3.2.2. Using Figures and Tables in the Report

As the saying goes, “A picture is worth a thousand words.” Adding figures and tables can strengthen the point you’re making and improve the quality of your report. A typical pitfall is that the writer includes a lot of figures to show all the analysis that they did. Keep in mind, the point of the report is not to list all

the analysis that you have done, but rather emphasize key findings and your process to find them. Therefore, sometimes less is more. That is, include only the figures and tables that play a key role in strengthening the arguments you are making.

Some things to consider before adding a figure or a table are:

- Does the figure strengthen my argument or provide additional insight?
 - If so, ensure that the figure is properly cited in the discussion of the report.
- What is the best way to present this finding?
 - Use a bar chart, pie chart, box and whisker plot, table, etc.
- Is the formatting consistent with other figures in the report?
 - Naming convention, font, borders, color, and spacing should be consistent with other outputs in the report.
 - Labels for tables always go on top. Labels for figures always go on bottom
- What is the appropriate section for the figure?
 - For some figures, the Appendix would be a more appropriate place for them rather than adding them directly in the report.

When deciding what figures to include, keep in mind that the goal is to either strengthen the argument you're making or provide interesting insights that you observed. When adding figures to the report, ensure that the figures are presented in such a way that the flow of the discussion is not interrupted. If figures or tables are utilized correctly, reading the report becomes easier and more enjoyable.

3.2.2. A. Figures and Tables Examples

A key to effectively using figures and tables is to keep a consistent theme throughout your report. In other words, for figures, choose which program you will use to make your figures, and ONLY use that program (e.g., Excel). For tables, pick a style within Word (or whatever processing program you use) and ONLY use that style for all tables. This adds to the overall professionalism of the report, and it's easier for you audience if they only have to orient themselves to a single visualization style.

Figures should be easily readable, with axis labels where necessary. The figure label should briefly explain to the audience what they are looking at. The example Figure 3 to the right is easy to interpret with the information provided, and the label gives further context.

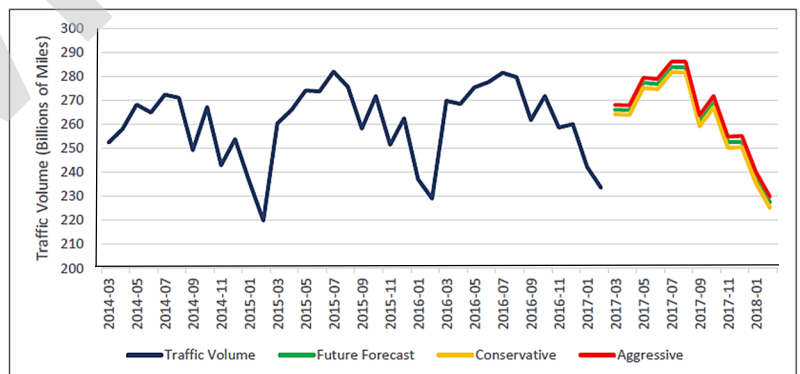


Figure 3: Example of displaying a future forecast.

Tables should also be easily readable. When showing numerical data, decide how many decimal points are needed to actually add value. Does a billion-dollar company care about '\$4.5678 Million', '\$4,567,821.23', or does '\$4.6 Million' give them enough critical information?

Tables are also effective for summarizing model diagnostics (Table 2) & error measurements (Table 3). The model diagnostics can be briefly explained in the narrative, while test types, error measurements and

results can be mentioned in footnotes. In Table 2, all tests would be given a footnote to define them (unless they have been defined earlier in the report). For simplicity, only the first test has been given a footnote to serve as an example.

Table 2 - Model diagnostic test, results, and conclusions

| | Test | Result | Conclusion |
|--------------------------------|----------------------------|-------------------|------------|
| Normality | Shapiro-Wilk ¹ | P = 0.81 | Pass |
| Independence | Durbin-Watson ² | DW = 2.04 | Pass |
| Homoscedasticity | Residual Plot | No shape detected | Pass |
| Linearity | Pure-error | P = 0.01 | Pass |
| Multicollinearity ³ | VIF | No VIF > 5 | Pass |

Table 3 – Model Error Comparison for Selected Models

| | ARIMA Best Fit | Winters' Method |
|------|-------------------|--------------------|
| MSE | 421.99 | 15.82 |
| MAD | 17.71 | 3.08 |
| MAPE | 11.24 | 2.52 |

As can be seen in both Table 2 and 3, the data is summarized efficiently, while also reducing the chance for confusion that comes from explaining these results in the narrative.

Another key take-away from this is the use of footnotes. Footnotes 1-3 below add succinct details for Table 2, without cluttering the narrative. The reader is given the option to

read the footnotes, allowing them to effectively move through the report. Use footnotes for details that would be cumbersome to search through the appendix for.

4. Report Writing

Writing an effective report involves preparation and practice. One thing to always keep in mind is to use common English terms when describing your methods and findings. Understanding your audience is essential, and broadening the intended audience is often beneficial. You don't want to assume that your reader has statistical or data analytical skills. The body of the paragraph should be a concise and clear representation of your research or analysis, which anybody should understand. If additional explanation is required, it should go in the appendix at the end of the report.

4.1. Structure

4.1.1. Abstract

This section should be written last and considered the "snapshot" of your report. It should provide a brief background of the industry/topic you are studying and a succinct summary of your findings. The Abstract is a standalone piece. The reader should be able to get the gist of your report from reading it.

¹ Shapiro-Wilk is a test to determine if residual values follow a normal distribution. A higher p-value (greater than 0.05) means the data is normally distributed (Tsai, 2018).

² Durbin-Watson is a test to detect auto-correlation using the lag of the residuals (DeLurgio, 1998).

³ Multicollinearity is not an assumption test. Rather, it is used to measure the relationship between x (independent) variables (Keller, 2017).

Specifically, you need to briefly and clearly state your problem (or goal) and highlight your most significant findings or recommendations with numerical support. Any findings listed here should also be mentioned in the introduction and in the conclusion. Also, the Abstract should coax the reader to read further. Novel findings with numerical support should be shared so that the reader gets excited about the subject matter.

4.1.2. Introduction

Many people find that the Abstract and the Introduction are extremely similar. They are, but the Introduction goes into greater detail about the contents of your report. In this section, you can elaborate more on the following: history and importance of the area of study, methods used and major findings, and an overview of the report. The writer can go as far as providing brief descriptions of the contents of each section.

The Introduction should capture the interest of your reader. It provides a first impression and allows the reader to understand your topic, why it is important, and how you plan to proceed with your discussion. Like the Abstract and Conclusion, the Introduction should be clear and concise and written for your reader — don't start your Introduction by overloading the reader with statistical data and analysis. Instead, begin with some relevant background information.

At the beginning of this section, you should provide some background for the report. Clearly state and elaborate on the purpose and motivations of the report. Readers should know exactly what problem(s) you are addressing. Also, it's important to show why readers should care about the problem(s). There should be a smooth transition between background and problem.

You can briefly (not in detail) introduce the methods you used in this report to solve the problems and provide the key findings at the end. When outlining the methods and findings, *DO NOT* mention all methods used or present findings that are insignificant. Highlight the key methods used and findings that impact the area of study the most.

Before ending this section, briefly state your insightful findings and valuable recommendations with some numerical supports. These should, at a minimum, consist of those listed in your Abstract. Additional input may be provided here that wasn't included before, as long as it still has an important impact on the study. These findings should be supported by concrete decision making (i.e. they must make business sense).

At the end of this section, you can list out the structure of your report and briefly state the function of each section. For example:

"This report is organized into four main sections (including the introduction). Section 2 explores the characteristics of the energy usage data in Ghausi Hall from January 2015 until May 2016. Section 3 presents the models that we developed to forecast building energy use for each energy source and our calculations of energy and financial savings for 2016. Section 4 provides a summary of our findings and recommendations for future exploration and analysis. The appendix provides the detailed statistical analysis of the methods presented in this report."

4.1.3. Data Characteristics

Don't underestimate the importance of this section — this is not an easy section to write. It is meant to introduce your data sets and present significant information. You should have performed an in-depth exploration of the data, but you should only include the relevant findings your report. These

characteristics should support the decision making behind your model selection. Don't forget, you're telling a story with your report — put yourself in your reader's shoes when writing the report.

Start with the basic, descriptive statistics that help the reader understand the data. These could be as simple as the mean, median, maximum, minimum, range, etc., as long as the information adds value to the story. Remember, include only the important phenomena⁴.

If your data set is purely time series data, you should discuss any significant features of the data set. (e.g. trends, seasonal, cyclical, outliers). Macro and micro perspectives are important in this process — don't forget to consider the economic implications of your data. For example, if you're looking at financial data around 2008, don't just talk about the downward trend in statistical terms. You should incorporate that the downward trend reflects the Global Financial Crisis that occurred during this period. Also, when looking at time-series data, it might make sense to look at incremental change. That is, graph the difference of $y_t - y_{t-1}$ and see if any significant patterns are observed. This analysis may lead you to believe that a transformation would be necessary given the construct of your data. If so, the transformation should be explored in the Model Selection section.

When you have time-series data with multiple variables, a good way to present this information effectively is to incorporate the data in a single time-series plot. As seen in Figure 4, time-series data is presented for four variables. This way, the reader can easily observe any interesting aspects of the data for all variables such as trend, seasonality, etc. Also, an effective way to present time-series data with different scales is to include them in one figure with different axes. As an example, Figure 5 includes time-series data for both variables: daily chilled water usage and cooling degree hours. Both variables have different scales and the information is presented in one figure.

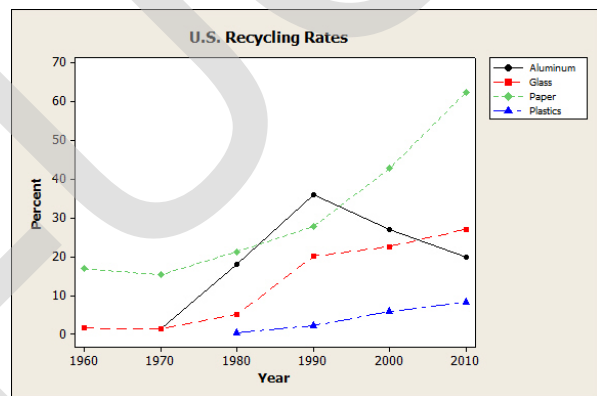


Figure 4: Time Series plot of Multiple Variables

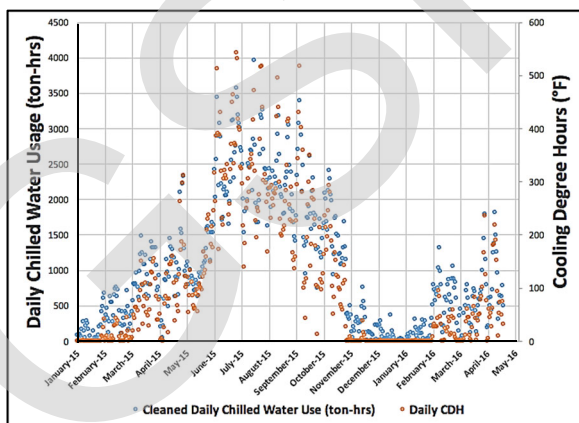


Figure 5: Time Series Plot with Different Scales

If you have time-series data with dependent and independent variables, you also need to analyze each variable before going any further and find out if there are any special features. Analyze the relationship between dependent variable and each independent variable. Consider using regression and find the relationships between different variables. Presenting one nice figure with a short summary is often the best way to present the data to the reader. Keep in mind that these are just suggestions. You should analyze your data sets thoroughly. This will help you decide which model(s) to use.

⁴ Consider using box plots — generally these are a concise, easily digestible way of describing the data.

Finishing your analysis, you should decide what to put into the report. In general, you should only include interesting insights about your data set that ultimately will drive the model selection. Simple plots and tables are useful, but remember to not overwhelm the reader. At the end of this section, it is often useful to include “Key Trends” which summarize the two or three significant observations from the data. At this point, do not refer to a statistics model, which will be discussed in the following section. Note that you need a smooth transition to the next section.

4.1.4 Model Selection, Evaluation and Interpretation

This section should convince the reader that you have chosen a decent model to forecast future data. This section is difficult to write and can be challenging to write in simple, non-technical, language. It is critical that language is simple and that you evaluate model(s) logically so you can convince the reader on the validity of your model before transitioning to future forecasting.

4.1.4. A. Model Selection

The first part of this section should be devoted to *why* you have decided to evaluate certain forecasting models. This discussion should tie in aspects of the data characteristics section to logically settle on appropriate models that can accurately fit and forecast the data given.

4.1.4. B. Model Evaluation

The second part, and the bulk of this section, is how you evaluated each model. There are three ways in which to evaluate a model: model fit, internal forecasting and diagnostics.

4.1.4. B.1. Model Fit

Evaluating model fit is simply testing to observe how well your statistical model “fits” against the actual data. The temptation in this section is to explain the “math” that went into developing the model; however, doing so will likely just confuse your reader. Simply state what model(s) were fitted to the data.⁵ After fitting the models, it is important to show the reader how the models performed with respect to fit (Figure 6). Use error measurements to quantify how well each model performed fitted against the data. Use MSE⁶ or MAPE⁷ to describe how “off” your model was from the data. Use ME⁸ to describe whether your model overfits or underfits. Try to stay consistent in the use of error measurements (i.e., do not report the MAPE for one model and the MSE for the other – this is comparing apples to oranges). Also try and limit the amount of error measurements used; the more you use the

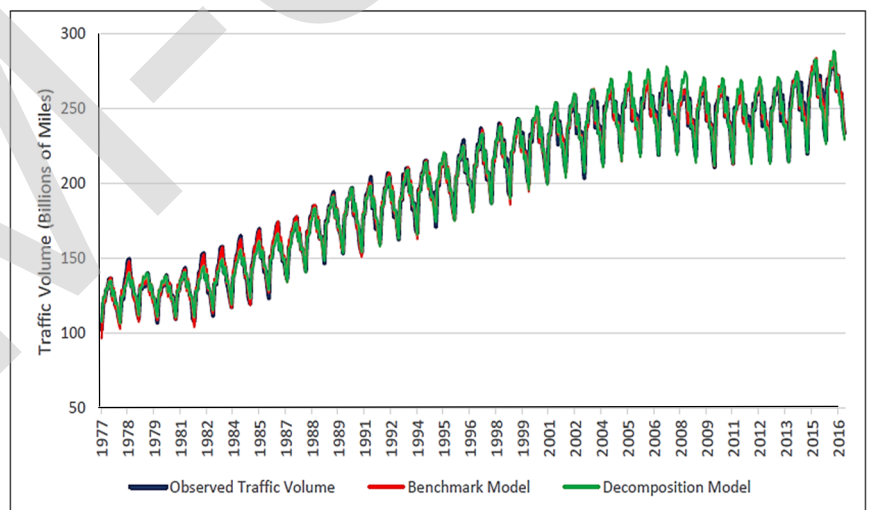


Figure 6: Example of two models fit against actual data

⁵ You can use the full data set to fit your models and then use the reduced data set for internal forecasting.

⁶ Mean Squared Error is the sum of the squares of each residual (DeLurgio, 1998).

⁷ Mean Absolute Percent Error is the error between your model and the data as a percentage of error (DeLurgio, 1998).

⁸ Mean Error is the sum of each residual (DeLurgio, 1998).

more technical the report becomes. You can end this part by stating which model fit best (i.e., which model had the lowest error when fitted against the data), and that you are going to conduct internal forecasting on that model.

4.1.4. B.2. Internal Forecasting

The next logical step in evaluating models is internal forecasting. If model fit tells you how well a model explains the data, internal forecasting tells you how well a model can forecast into future periods. Prepare a reduced data set⁹, fit your model to that reduced data set then forecast for the time periods you reserved. Just with the model fit, a visual inspection (Figure 7) as well as error measurements of the internally forecasted section should be conveyed to the reader. Additionally, display a confidence interval¹⁰ for your forecast against the actual data to get a sense of whether your conservative, expected or aggressive forecast is closest to the actual data (Figure 7)¹¹. You can end this part by stating which model's internal forecast fit best (i.e., which model's internal forecast had the lowest error when forecasted against the actual data using the reduced data set to model).

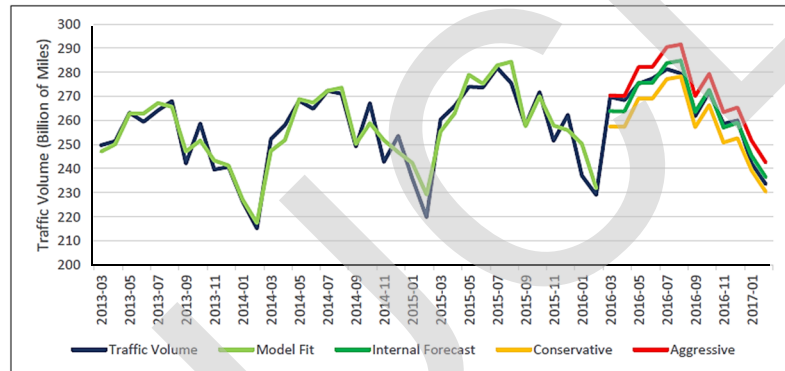


Figure 7: Example of a models internal forecast fitted against actual data with its confidence interval

4.1.4. B.3. Model Diagnostics

The final tool used to evaluate a model is diagnostics. Diagnostics for your model should represent the model fitted to the entire data set (as opposed to the reduced data set used for internal forecasting). A model's ability to stand up to diagnostics will speak to each models' integrity for explaining and forecasting future data. Just as with the model fit section, there is a temptation to get technical here. Even though you will run informal and formal tests for normality, heteroskedasticity, autocorrelation, etc., it is important to just state any significant findings here.¹² For example, in an ideal situation, a model that is being evaluated passes all diagnostics. In this situation, it is important to simply state that this model is validated for future forecasting because it has passed all diagnostics. In a less ideal situation, a model will fail certain diagnostics. In these situations, you can state to your reader which test the model(s) failed and, most importantly, how the failure(s) effect potential future forecasted values or any changes to your model(s).¹³

⁹ Set aside the last 5-10% of your data set

¹⁰ The Residual Standard Error from a model can be used to calculate confidence interval for forecasted values

¹¹ In the event that either your conservative or aggressive internal forecast mirrors the actual data better than the expected value, then whichever band is closest to the actual should be used for future forecasting as your new "expected values." The implied task in this case is to then create a new confidence interval around the new expected band. For example: during internal forecasting, the aggressive band most closely follows the actual data; therefore, when you conduct future forecasting with the full data set, the aggressive future forecast now becomes your "expected" values, and you may create a new confidence interval around this band.

¹² All the "math," figures and charts from diagnostics should get its own section in the appendix.

¹³ If a model is selected for future forecasting that failed some critical assumptions, an increase in the range of future forecasted confidence intervals may be appropriate.

4.1.4. C. Interpretations

The last part of this section is bringing everything together for the reader. If you have been comparing two models, this part provides an opportunity to recap how the models performed compared to each other. Ideally, you will have a front-runner model that clearly outperforms others in model fit, internal forecasting, and diagnostics. If you have a situation where there are mixed results, then this part of the report should be devoted to a discussion as to why you will pick one model over the other.

Prior to transitioning to the future forecasting section, this part gives you an opportunity to share with your reader any peculiarities about your selected model. For example, if you are using a time series model (decomposition, Winters, etc.), how well did the model capture the seasonality or cyclical effects observed in the actual data? If you used regression, a discussion about your selection of independent variables and their resulting coefficients (or partial correlation to other variables) may be appropriate. Additionally, if you are using the benchmark method or inverse forecasting method, it may be prudent to state what critical assumptions you made here (see Appendix B for an example with benchmark forecasting). Basically, if it is important for your reader to understand something about your model (including any limitations) prior to applying it for future forecasting, then it should be stated here. This part gives you a great opportunity to relate the statistics of your model and analyses to your overall business application.

This section should end with a clear statement of which model will be used for future forecasting and an appropriate transition to the next section.

4.1.5 Future Forecasting

With a model selected, you can now transition on the critical component of your report: future forecasting. The method for future forecasting should follow the same approach you took with your chosen model when you performed internal forecasting. However, you will need to construct your model using the full data set and apply the same assumptions.¹⁴ A critical aspect of future forecasting is to be cognizant of compounding error, in other words, the further into the future forecast, the more wrong you will likely be. A good rule of thumb to apply is to forecast one “period” into the future. For example, if you are working with data that is seasonal month to month, then forecasting one year into the future would be prudent.

Your future forecast should be presented as a confidence interval representing conservative, expected and aggressive scenarios. It is important to recall from internal forecasting which of these bands (conservative, expected or aggressive) most followed the actual data. Looking at Figure 7 on the preceding pages, the expected forecast in this scenario most follows the actual data. Therefore, I would indicate to the reader that the “expected band” of values in my future forecast is likely what will materialize in the future. Again, displaying your future forecast can be a powerful tool to show the reader that you have captured the trend, seasonal and cyclicity of the data. See Figure 8 for a continuation of the example used in Figure 7.

¹⁴ If you use the benchmark method with multiple linear regression, your benchmarks should remain the same as when you conducted internal forecasting.

It is important to note that error measurements are not used in analyzing your future forecast (because there is no actual data to compare it to). One way to informally analyze your future forecast, is to graphically display your forecast alongside the latest data (Figure 8). If your model's forecast captures the overall "shape" of the data moving forward, it will make it easier to convince your reader that your forecast is believable. If your forecast does not follow the latest data's "shape," then you should convince your reader why the forecast looks differently.

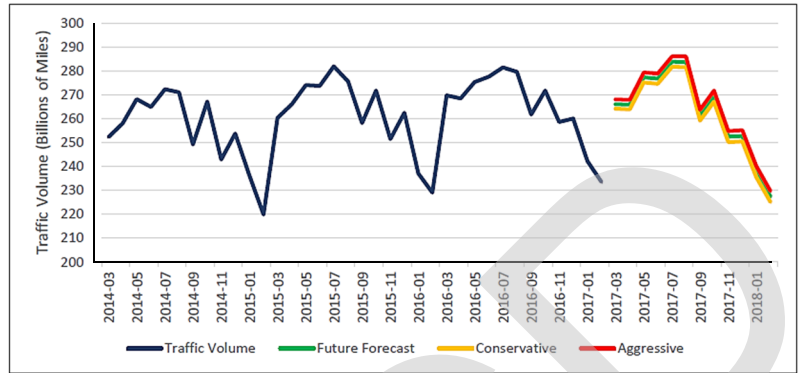


Figure 8: Example of displaying a future forecast.

4.1.5. A. Multiple Linear Regression Considerations for Future Forecasting

Using multiple linear regression (versus a strict time series model) has unique challenges when conducting forecasting. The following are points advising the use of regression in forecasting.

4.1.5. A.1. Forecasting via Regression with Variables that are not time related

When you would like to forecast, but the variables you are working with are not time related, the two possible approaches to use are benchmark forecasting and inverse forecasting. Suppose you have a dataset of weight and height of some population. You would like to use the height to forecast the weight of the individuals. You can develop a benchmark such as mean, median, mode of the height. When deciding a benchmark, it is important to justify why you're using the specific benchmark given the context of your problem. Next, using the benchmark, you can forecast the weight.

Another approach is to use inverse forecasting. Suppose you have a dataset of revenue and number of employees for fifty different companies. If you have a specific revenue target in mind, you can inverse forecast the number of employees to achieve that revenue target. That is, in general terms, given a forecasted value of a dependent variable, you can inverse forecast the level of a controlled independent variable to achieve the forecasted value of the dependent variable.

4.1.5. A.2. Forecasting via Regression with Time Series Data

When working with regression and you have time series variables, benchmark forecasting (e.g., see Appendix B) and inverse forecasting can be applied in this context as well. Another approach is to first forecast each independent variable. Next, use the generated forecasts for each independent variable to forecast the desired dependent variable. Please note that this may lead to compounded error. That is, since you're using forecasts that have error already embedded in them and use them to forecast the dependent variable that will only compound the error measures.

4.1.6 Conclusions and Recommendations

Like most movies, the ending is the most interesting and important part. In this section, you tie the rest of the paper together. Summarize your findings, include your comments, suggestions/recommendations, and the impact of these findings for the future of the company and future studies. These are what your audience wants to learn from your report. Link findings to the problems you stated in the Introduction, and make strong recommendations concerning the appropriate business division, whether it is finance, marketing, operations, etc.

Note: It is critical you link the findings first mentioned in the Abstract and Introduction to what is presented in the Conclusion. This is another reason it is advisable to write the Abstract last.

First, the conclusion needs to draw from your model and forecasting. For instance, your model and forecasting suggest that revenue can increase 20% if the company allocates an additional \$10 million to marketing. You need to provide numbers from previous sections to draw this conclusion. Also, different scenarios should be considered. And based on your analysis, explain which scenario is most likely.

Next, you need to provide recommendations and an execution plan. Remember, the most successful reports provide a menu of recommendations, not just one. It is fine to push a particular recommendation, but be cognizant of your readers. Formulate suggestions for different types of readers. A good example of this is formulating different investment recommendations based on an investor's level or risk aversion. Back to the example above, even though a 20% increase in revenue is attractive, whether the company can increase marketing budget by \$10 million is questionable. You should provide the potential source of funding and brief marketing budget layout, opening a door for the audience to think about. Finally, your recommendations should be actionable. Any non-actionable recommendation is meaningless to your company.

Consider limitations of your analysis and why they presented themselves (time, money, etc.). Also, describe their possible remedies (more advanced techniques, etc.). Make sure any mention of limitations is subtle and not overemphasized. Do emphasize the impact to the business. Be sure to end your report strong; the reader should walk away feeling that they have learned something new, unique and useful. A well-written conclusion provides a differentiating factor from other reports.

4.1.7 Appendix

Most graphs and tables that haven't been mentioned in the previous sections of this guide should be included in the appendix, unless they are absolutely necessary in the body. Also, all complex statistical analyses MUST go in the appendix. Leaving these in the body of the report will add clutter and draw the readers' focus away from your message. Be sure to label exhibits clearly and appropriately. Also, add brief captions to exhibits to explain the contents of each. These appendices must be able to stand alone. They can support claims in your main report, but a reader must be able to look at appendices and understand them without having to read the rest of your report.

The exhibit numbers of each graph should correspond to its sequence within the Appendix. In addition, the caption should cite the section it is referenced in. However, anything you don't cite in the body content, DON'T include in the appendix. The appendix is for complimentary/supporting material and should not be treated as a dumping ground for every graph and table your software package produced.

In the beginning of the Appendix, please describe the contents in the Appendix such as Appendix A. Error Measurements, Appendix B. Alternative Models, etc.

4.1.8 References

All references should be cited and listed in this section, choose a format (MLA, APA, Chicago, etc.) and stick to it. Articles, journals, and books should be referenced directly in the narrative (e.g., Smith, 2012), and websites should be listed as a footnote¹⁵.

Be sure that all references are cited in the body of the report. A reference page is worthless if the reader doesn't know which article informed which section. A possible format of references is given on page 26.

4.2 Miscellaneous Tips

Transitions are very important between sections. They help the reader follow your train of thought. This seems trivial, but it is actually critical in professional report writing.

In the body of the report, it is advisable to limit the empty space surrounding graphs or tables. To do this, use the "Text Wrapping" function in Microsoft Word or similar word processor. For reference, there are several sections within this handbook of figures with text wrapping.

In addition, a business report template is provided (Appendix C). This format can be used to create a layout for your own business reports. Use this layout with the sample report provided (Appendix D) to help guide your writing.

5 Post-Report Writing

Your work doesn't stop when the report is done. Recall that we recommended you get some feedback on your first draft. Also, PROOFREAD before you submit the report. Below are two additional recommendations.

5.1 Submission Etiquette

If you are awaiting an interview, the proper time to submit an example of your work through a business report is at least 24 hours before the scheduled interview through email. This provides enough time for your interviewer to look at your report, and it will allow you to discuss it during the interview. It is generally advisable to not submit a printed copy unless explicitly asked to.

As an accompaniment to a presentation, you can print out your business report in advance and circulate it to your audience just before presenting. In this case, emailing the report in advance can result in too many premature questions and interrupt the flow of your presentation. Handing out the written report and sending a copy through email at the end of your presentation, allows your audience to continue reading your work if they are so inclined, which extends the life of your scheduled presentation and leaves a strong impression on your team.

5.2 Second Report / Report Modification

As mentioned in the Conclusion and Recommendation section, your analysis might have some limitations. If you have a second interview or more resources available, you can modify your report accordingly. Remember, you still need to consider your purpose and audience when you modify the report. Also, the

¹⁵ www.reference.com, accessed on May 20, 2018

second report should be significantly better than the first one in terms of accuracy, conclusions, or recommendations.

6 Conclusion

As stated at the beginning of this handbook, professional written communication skills are essential in the business world. In the foreseeable future, these skills will only be more important as the Big Data era has arrived. Big Data generates more inclusive data. You will find that asking the right questions and translating data into action will be a key skill, especially for MBA students. Communicating insights to decision makers and creating actionable recommendations can differentiate you from others.

As you go through this process, we suggest you step back and look at the big picture. Your report, findings, and recommendations should align with your company's vision. More importantly, like Prof. Tsai once said, "life is the same as writing a good report. You face challenges you've never seen before, you overcome them with persistence, and that will make you stronger."

Last suggestion: Show me your guts!

Appendices

There are three appendices in this handbook:

- Appendix A is a list of helpful data sources
- Appendix B is an example of benchmark forecasting
- Appendix C is a report template for quick reference
- Appendix D is a sample business report for your reference

Appendix A: Data Resources

The attached supplementary material provides some useful links for finding data sources. They are up-to-date as of this updated handbook's publishing in May 2021, and please be advised that some of them might not be valid as you're reading this handbook.

Appendix B: Example of Benchmark Forecasting

Benchmark forecasting is a powerful tool that can be difficult to understand and even harder to describe using non-technical terms. Below is an example to help guide your framing of using this statistical method within a report.

Start with your structure for your multiple linear regression model:

$$\begin{aligned} \widehat{TrafficVolume}_i = & -1.65 + 0.86 * Lag1TrafficVolume - 0.0014 * GasPriceIndex + 1.59E - 8 \\ & * U.S.WorkingPopulation + 1.92E - 7 * TotalAutoSales - 0.013 \\ & * LTGovernmentBondYield + 0.0098 * CPI - 0.0041 * TimeIndex + 1.31 * Mar + 0.41 \\ & * Feb ... - 0.074 * Jan \end{aligned}$$

Here, we have the dependent variable, *traffic volume*, being explained by 7 independent variables (lag1 traffic volume, gas price index, population, auto sales, bond yields, CPI, and time) and 11 dummy variables for months. We simply need to input our benchmark values for forecasting.

The next step in forecasting using the benchmark method is to select the appropriate benchmark for each independent variable in your model. Benchmarks should be chosen based on what you believe to be representative of that variable during the forecasting period. Simple benchmarks include using the mean or median value for that variable in the forecast. Another example could be to apply a simple growth rate to a variable during the forecasting period. Using the model structure previously stated, below are the benchmarks used (with rationale in parentheses):

- Lag1 Traffic Volume: the previous month's traffic volume
- Gas Price Index: 257.78 (average of last 60 observations – gas prices are cyclical: seems appropriate to use the average of the last 5 years)
- Working Population: 205,903,000 (max value of the data set – used max due to slow/steady growth)
- Auto Sales: 707,086 (average of all observations – highly cyclical data: average seems appropriate given movement around the mean)
- Long term bond yields: 2.03 (average of last 24 observations – interest rates are cyclical: used average of most recent interest rates)
- CPI: 243.60 (max value of the data set – same rationale as population: slow growth)

- Time index: insert the time index for the forecasted value
- Month: insert 1 for the coefficient on the dummy variable for the corresponding month

Applying benchmarks for forecasting (using March, 2017 [time index = 482], as an example):¹⁶

$$\begin{aligned} \widehat{\text{TrafficVolume}}_{\text{March},2017} &= -1.65 + 0.86 * (15.28) - 0.0014 * (257.78) + 1.59E - 8 * (205,903,000) + 1.92E - 7 \\ &\quad * (707,086) - 0.013 * (2.03) + 0.0098 * (243.60) - 0.0041 * (482) + 1.31 * (1) \end{aligned}$$

$$\widehat{\text{TrafficVolume}}_{\text{March},2017} = 16.31 \rightarrow 16.31^2 = 266.03 \text{ Billions of Miles}$$

Appendix C: Report Template

On the following pages, you will find a template that will help you organize your own business report. If you have an electronic version of this document, feel free to copy and paste into your word processor, but please be sure you read over your finalized report to ensure the template is entirely “filled in”, and no sections of the template remain incomplete. There are few things worse than typos on a resume or cover letter — this also applies to your business reports.

¹⁶ In this example, the dependent variable was transformed using a square root transformation (to control for normality issues). This means that the model’s output needs to be squared to get the actual forecasted value.

Business Report Writing Handbook
Report Template

[Insert Relevant Graphic Here]

Title of Report

Date

Name

Name (additional authors as appropriate)

Table of Contents¹⁷

| | |
|--|-----------|
| Abstract | 24 |
| 1. Introduction | 24 |
| 2. Data Characteristics | 24 |
| 3. Model Selection, Evaluation and Interpretation | 24 |
| 3.1. Model Selection | 24 |
| 3.2. Evaluation | 24 |
| 3.3. Interpretation | 24 |
| 4. Forecasting | 24 |
| 5. Conclusions and Recommendations | 24 |
| Appendices | 24 |
| References | 24 |

¹⁷ For your actual Table of Contents, use Word's Table of Contents feature, which will allow you to auto-update headers and page numbers as you write your report.

Abstract

For guidance on writing this section, refer to page 9 of this handbook.

1. Introduction

For guidance on writing this section, refer to page 10 of this handbook.

2. Data Characteristics

For guidance on writing this section, refer to page 10 of this handbook.

3. Model Selection, Evaluation and Interpretation

For guidance on writing this section, refer to pages 12-14 of this handbook.

3.1. Model Selection

For guidance on writing this section, refer to page 12 of this handbook.

3.2. Evaluation

For guidance on writing this section, refer to page 12-13 of this handbook.

3.2.1 – Model Fit

3.2.2 – Internal Forecasting

3.2.3 – Model Diagnostics

3.3. Interpretation

For guidance on writing this section, refer to page 13-14 of this handbook.

4. Forecasting

For guidance on writing this section, refer to page 14 of this handbook.

5. Conclusions and Recommendations

For guidance on writing this section, refer to pages 15-16 of this handbook.

Appendices

For guidance on writing this section, refer to page 16 of this handbook.

References

For guidance on writing this section, refer to page 16-17 of this handbook.

Appendix D: Sample Business Report

The below business report was a final project for MGT 285 that analyzed the impact of energy efficiency updates made at UC Davis. To access the report, please click on the below link.

Business Report Link: <https://gsm.ucdavis.edu/profile/chi-hing-tsai>

References

- Beristain-Barajas, R., Havens, K., & Perez, J. (2015). *Droughtproofing: Is Desalination the Answer to California's Water Problem?* Business Report.
- CollegeBoard. (2004). *Writing: A Ticket to Work...Or a Ticket Out A Survey of Business Leaders*.
- DeLurgio, S. A. (1998). *Forecasting Principles and Applications*. New York, NY: Irwin and McGraw-Hill
- Fried, J. (2010). *Rework*. Crown Business.
- GMAT. (2017, July 12). *Employers Seek Communication Skills in New Hires*. Retrieved from The Official GMAT Blog: <https://www.mba.com/us/the-gmat-blog-hub/the-official-gmat-blog/2017/july/employers-seek-communications-skills.aspx>
- Keller, G. (2017). *Statistics for Management and Economics*. 11th ed. Boston, MA: Cengage Learning
- NACE. (2016, April 10). *Job Outlook 2016*. Bethlehem: National Association of Colleges and Employers (NACE). Retrieved from Naceweb.org: <http://www.naceweb.org/s11182015/employers-look-for-in-new-hires.aspx>
- NYU Data Services. (n.d.). *Which Statistical Software to use?* Retrieved from Statistical Software Guide: <https://sites.google.com/a/nyu.edu/statistical-software-guide/summary>
- Tsai, C. L (2018). *Notes of Forecasting and Managerial Research Method*. Graduate School of Management, UC Davis

7. Recommendations for Instructing a Course Utilizing Data-Driven Business Reports

As an instructor seeking to integrate data-driven business reports into a curriculum, there are three areas that require close attention to ensure the students derive maximum value from the course. Prior to the course, students need to be fully aware of the expectations of data-driven business reports. On the first day of class, students are recommended to form groups and individual team leadership should be assigned. Additionally, a comprehensive overview of the course should be re-emphasized. Finally, throughout the course, Readers/TAs and the instructor should work in tandem to mark the progress of and provide assistance to individual teams, when necessary.

7.1 Prior to First Day of Instruction

It is recommended that the instructor post the following documents on their respective student-facing platform prior to beginning of instruction:

1. A comprehensive syllabus (with a link to the following three documents)
2. Sample homework reports and final project reports
3. A “Teamwork Meeting Guidelines” document
4. The “Handbook for Writing Data-Driven Business Reports”

7.1.1 The Syllabus

It is recommended that the instructor clearly outlines the intent of this course in the syllabus. It does not matter the context of the course, be it finance, marketing, strategy etc., data-driven business reports can be integrated into various fields of study.

To ensure a successful student and instructor experience, the instructor should clearly communicate the purpose and motivation of the course in the syllabus. The overarching theme of the purpose and motivation statement is meant to demonstrate to the students that the course is designed for them to learn the course material while also being able to communicate it clearly and formally to the general population. Neither aspect is recommended to be emphasized as more important than the other. Students should understand that presenting the material in an easy to understand and usable format is just as important, in a professional setting, as the complex analysis performed on the data.

The syllabus should be distributed to the incoming students well-prior to the first day of instruction. This allows students time to fully recognize the expectations of the course, and how the course material will be integrated into the assigned reports. It is recommended that the following three documents be linked within the syllabus so that students have an even better understanding of the expectations.

7.1.2 Sample Reports

A link that gives students access to sample homework and final project reports, created by previous students that have completed the course, is recommended to be provided in the syllabus. These sample reports provide students with a clear visual representation of report expectations, both from a content and formatting perspective. The samples can also be essential in creating efficiencies for both the instructor and the Reader/TA. Limited in-class time should be spent on instructing students of the exact expectations of a well-done report, allowing the instructor to maximize the material being covered. With

grading and advising, the Reader/TA can immediately provide detailed feedback to the students without needing to dedicate time to report-writing basics. If this is an instructor's first time teaching this course, and there are no sample reports from previous classes in the field of study he/she is teaching, sample reports are available at <https://gsm.ucdavis.edu/faculty/chih-ling-tsai>. If these reports do not fit the specific needs of the course, it is recommended the instructor uses the prevailing academic journal format of his/her area of study.

7.1.3 Teamwork Meeting Guidelines

It is recommended that the instructor prepare a "Teamwork Meeting Guidelines" (e.g., section 7.5.1) document to assist teams in how to properly allocate tasks among group members and best utilize time spent during the process of generating the report. The report writing process is very time consuming, and a standardized workflow is beneficial for the students that are new to the process. This document will provide groups with certain milestones that should be completed in the report writing process, and in what stage of the process that the milestones should be completed. A link to this document should be provided in the syllabus.

7.1.4 Handbook for Writing Data-Driven Business Reports

A link to the above handbook is recommended to be provided in the syllabus.

7.2 On the First Day of Instruction

The first day of the course is crucial to the success of students and the value that each derives from the course. It is recommended the instructor complete the following tasks:

- Students should form their groups. The suggested group size is no greater than 3 or 4 students. It is recommended that one of the students is self-nominated as the "Team Leader".
 - Responsibilities of Team Leader: The Team Leader will assist the professor in tracking the team's progress. He/She is meant to serve as an entity to keep the team cohesive and ensure that each member is meeting his/her personal milestones. The Team Leader should also be the individual responsible for collecting and submitting each weekly report and the final project. In the case of any challenges, Team Leaders are expected to notify the instructor promptly.
- The instructor should set up a time to meet each team, privately, within 5 days of the class beginning. During this meeting, Team Leader needs to present their homework timeline for each week and the associated workflow plan, using the "Teamwork Meeting Guidelines" for reference. Key questions that are recommended to be asked: How are they going to accomplish their goals? How will they allocate work?
- The instructor should emphasize the use of the "Handbook for Writing Data-Driven Business Reports." Students should be familiarized with the guide and understand that it is expected to be used throughout the course.
- Each student is expected to submit a summary sheet of the material each week. The purpose of the summary is to gauge whether each student understands the material, on an individual level, and helps to prevent students from underperforming within their groups.

- Students should understand they are expected to submit a 2-3 page final project proposal sometime in the middle of the quarter/semester. They should begin thinking about this project immediately, in regard to the collection of data, purpose, motivation, possible models, etc. A formal meeting should be scheduled at the earliest convenience to discuss any challenges in making a final proposal.

7.3 Tracking Progress

Readers and TAs should be aware of the importance their role has in student success during this course. The feedback given to students on their assignments shapes their understanding of the course materials and guides their report writing for all subsequent reports. It is recommended that the instructor monitors grading closely to aid teams when necessary. The following highlights the steps that a Reader/TA should undertake when grading/critiquing an individual report, and how best for the instructor to monitor the Reader/TA to ensure consistency and accuracy. The Reader/TA's grades should serve as a recommendation for the instructor. The instructor should ultimately be determining the final marks for each assignment.

7.3.1 Reader

7.3.1.A Before Grading

The Reader/TA should thoroughly review the “Handbook on Writing Data-Driven Business Reports.” There also needs to be enough time given to students, between reports, to digest comments from the previous assignment before submitting their next assignment. Thus, the recommended turnaround time for the grading process is 3 to 4 days.

7.3.1.A.1 Grading Philosophy

The Reader/TA is here to help the students improve on their data-driven report writing abilities and to encourage fairness. He/She should be guiding the students. Each comment given on a report should inform and offer council. Simply saying “Wrong!” and subtracting points does not add value to the student's experience. The Reader/TA should track, and grade based on progress made by each individual group, while also grading relative to other groups. Some groups will be inherently more skilled than others and should be graded as such. However, this should not inhibit any group from achieving high marks relative to their own perceived ability. On a week-to-week basis, it is important to reward students that are delivering exemplary reports with higher grades than underperforming groups. However, incremental improvements should also be rewarded, but to a smaller degree. This philosophy encourages fairness across all skill levels.

7.3.1.B Grading Clarity & Narrative Recommendations

- Each paper needs a clearly stated purpose and motivation (Who are you? Why does the problem need to be addressed now? What are the repercussions if the problem is not addressed?).
- Transitions between sections should be smooth. Clear through-line from problem to solution. The paper should read like the building of a pyramid (starts with a foundation and works its way up to the solution).

- Check tables/figures, style, language, references, and appendices (use the “Handbook on Writing Data-Driven Business Reports”).
- Students should be creating original tables/figures. Each table/figure should have a clear purpose and motivation. Technical information is expected to be summarized in easy to read tables/figures.
- Students need to use a forecasting model to address the problem (or any technical analysis that the in-class material is meant to provide).
- Always ask yourself about each sentence - does this bring value to the paper?
- Each report should offer three courses of action for the stakeholders: conservative, moderate, and aggressive.
- Any material in the appendix should be referenced to in the body. All material in the report should be relevant and insightful to the reader. The appendix should be precise and concise.
- Can grandma understand this? Will the non-technical stakeholder derive meaning from each statement/figure.

7.3.1.C Grading Technical Analysis Recommendations

- Critique the reasoning (are their decisions justified?). Students have the freedom of creativity in their technical analysis. It is the Reader’s job to decide if each of their decisions came with good/just reasoning.
- Address significant shortcomings (i.e. did not include an internal forecasting section). There are certain steps that should be followed by each group, regardless of which forecasting model each selects.
- Investigate outlandish claims or numbers (use your judgment!).

7.3.1.D Reader/TA General Feedback Form

- When grading reports, place specific feedback in the form of comments in the document margin. The comments should address specific areas of improvement and should also be reflected in the Reader/TA General Feedback Form.
- Compile individual group’s comments by section and organize Feedback Form to read section-by-section.
- The Feedback Form should open with a “general comments” section giving feedback on the report as a whole. Comments that occur repeatedly throughout the paper should be placed in this section.
- Emphasize significant comments in each section, especially statistical errors. Focus on overarching recommendations/issues.
- Highlight repeat comments and errors from the previous report. Track their progress, both positive and negative. If teams are repeatedly making the same errors, an increased amount of points should be deducted each time.
- The Feedback Form and the margin comments should work in concert to provide a total accounting of the team’s performance on the assignment. The comments should provide value and context to the report writer and offer advice for future reports.

7.3.2 Instructor

Once the reader has completed his/her grading duties, the instructor is recommended to review the “Reader/TA General Feedback” for each group. This document, created by the Reader/TA and highlighted in the above steps, is an efficient means for the instructor to ensure the grading is consistent and fair. The instructor should also be using this document to monitor teams’ individual progress. Again, it is recommended that the instructor be responsible for assigning final grades for assignments.

An additional role of the instructor in the feedback process is to provide the students with a “General Comments” document. This document is given to the entire class and highlights common errors in the technical analysis and report writing structure that was uncovered in multiple reports by the TA/Reader. This document should be provided to the team leaders, along with the “TA/Reader General Feedback Form,” within 5 days, to ensure the teams have necessary time to implement the changes into their subsequent reports.

If the instructor finds any teams are not meeting the established requirements after the first assignment, he/she should set an immediate appointment with the underperforming teams prior to the due date of the next report. Setting an early foundation of the report writing process and workflow is crucial to the value that each student derives from the class.

7.3.2.A Instructor Philosophy

This course is designed to be a very challenging 10-week¹⁸ training that will test each student’s mental endurance, time management skills, and technical acumen. The goal is to shape the students into future, ethical leaders that can prosper in difficult professional settings. It is a hope that this course has a profound impact on each student’s life well beyond the 10-week curriculum. Competition grows stronger as time passes, and technology is rapidly changing the landscape of the job-market. Creative, independent thinking is necessary to succeed in such a fast-paced world where information is so plentiful. Future leaders will require the skill to synthesize a multitude of information sources and subject matter into one useful and easy to understand product/deliverable.

7.4 Final Project Preparation and Presentation

7.4.1 Preparation

It is important for the instructor, throughout the quarter/semester, to ensure students are collecting data and they are finding an appropriate purpose and motivation for their team’s final project. From the first day of class, instructor should highlight the need to get started on this project early.

It is recommended that the instructor assign each group a two-page written final project proposal, due during the middle of the quarter/semester. After receiving the proposals, the instructor should set up a meeting with team members to review their proposals to ensure they are cohesive internally. It is also recommended for each group to have two separate proposals of potential projects. Often, students will find that their initial project choice does not satisfy the requirements of the project. Having two proposals can save each group time during the final weeks of the course, as opposed to actively searching for an alternative.

¹⁸ In the case of a semester length course (16 weeks), it is recommended to adjust the number of assignments and final project specifications to accommodate the longer course length.

7.4.2 Final Project Presentation

A successful project presentation will incorporate the following attributes:

- Presentation length should be 25-30min including Q&A. Each team member should be encouraged to present for an equal duration of time.
- Formal dress. Students should act as if it is a presentation to key stakeholders in a workplace setting.
- The presentation should be an executive summary of their report. Make concrete recommendations with numerical support. Presentation language should be in layman's terms. In the workplace, it is common that a manager will first present to stakeholders before distributing official reports for those that are interested in further learning about the project. This is meant to simulate this common workplace dynamic.
- Goal of project should be very clear to listeners. Presentations should flow from their initial goal statement and build towards a clear conclusion.
- Teams should rehearse extensively. No index cards should be allowed, and all team members should be familiar with the entirety of the project and be able to answer questions.
- As a component of the final project grade, it is recommended that a certain quantity of the total points should be directly related to the presentation. These grades can be given on either an individual or a team basis. This encourages individual ownership for each team member in regard to his/her grade.

7.5 Supplemental Materials

The following are sample materials that can be used by an instructor to help students establish a consistent workflow.

7.5.1 Teamwork Meeting Guidelines

Teamwork Meeting Guidelines

General Information:

The following is a suggested meeting schedule for four (or five) weekly assignments. By no means is this expected to be the optimal schedule for your team. Use this information as a starting point and adjust it according to your team's unique situation. The team leader should the instructor the team's guidelines with each member's name, signature, and date in a week from the first lecture.

It is suggested for the team to meet three times per week to produce the report. Meeting duration and deliverables are merely suggestions. Time spent in meetings will vary depending on the work accomplished in prior meetings as well as work completed outside of team meetings. If a team member misses class for any reason, it is his/her responsibility to learn the necessary material prior to attending the first meeting. This is so as to not derail the team's weekly progress. One team member's absence should not be the responsibility of all others on the team.

Weekly Meeting Schedule

Meeting #1: Brainstorming Session

Purpose: Brainstorm ideas for the assignment's narrative

Meeting Duration: ~1 hour

Deliverables for Next Meeting: Statistical analysis & narrative research

Overview:

Ideally, this meeting should happen as soon as possible after the team receives the assignment for the week. This should be a relatively short meeting where team members exchange initial ideas about the narrative as well as discuss the work to be done for the next meeting. The goal is to develop a general idea how the data set will be used and in what context. This general idea will drive the narrative development as well as the application of the statistical analysis.

Another important part of this meeting is that the team should decide the next steps. This would include having a general idea of what statistical models should be utilized as well as additional research for the narrative. For the next meeting, each team member should complete individual statistical analysis and narrative research.

Meeting #2: Statistical Analysis and Narrative Development

Purpose: Finalize the idea for the narrative and complete the statistical analysis

Meeting Duration: 2-4 hours

Deliverables for Next Meeting: completed rough draft of the assignment

Overview:

This is the first thorough team work session for the assignment. The first goal of the meeting is to finalize the narrative idea. Next, after reviewing the statistical analysis, the team should agree on what analysis would be presented in the assignment. The team must develop a clear understanding how the statistical analysis will enable them to create a full narrative with meaningful insights.

Once this is complete, the team must divide up each section of the report and assign them to a teammate. During this session, the team may have questions and need to seek assistance from Prof. Tsai. If this is the case, any questions should ideally be answered before the next meeting.

Meeting #3: Finalizing the Report

Purpose: Finalizing the report

Meeting Duration: 2-3 hours

Deliverables for Next Meeting: Team leader submits the report to the instructor

Overview:

This is the final work session for the weekly assignment. The team should have the rough draft of the report at this point. All sections of the assignment are completed (i.e. Executive Summary, Introduction, ... , Conclusion, Appendix, and References). The team reviews the full report ensuring that there is a clear flow. Also, the team makes sure that the report provides quantifiable insights that have a tangible impact on some aspect of the organization. The team should have this meeting at last 24 hours prior to the assignment submission deadline. This can give the team sufficient amount of time to do any final editing.

Meeting Guidelines for the Final Project

Please refer to the above Part and prepare your team's meeting Guidelines for the Final Project.

Additional Information:

Meeting Format:

- Ideally, in person. If not, video chat applications (e.g. Google Hangouts, Zoom, Skype, FaceTime)

Report Collaboration:

- **Google Drive**
 - Google Drive provides great collaboration tools.
 - Research and statistical analysis can be uploaded to Google Drive for easy sharing and collaboration.
- **Microsoft OneDrive**
 - Google Drive has limited formatting functionality compared to Microsoft Office files.
 - Final reports could be done in Microsoft OneDrive as they will require the use of Microsoft Word.
- **Box**
 - Box is an alternative to Google Drive and Microsoft OneDrive.
- **Dropbox**
 - Like Box, Dropbox is an alternative to Google Drive, Box and Microsoft OneDrive.



UC DAVIS

**GRADUATE SCHOOL
OF MANAGEMENT**

Data Resources Handbook

FOR CURRENT AND PAST STUDENTS OF MGT/P/B – 285

(PLEASE DO NOT CIRCULATE OF MAKE ANY COPY)

ISSUED: 05/10/2021

PREPARED BY: PRAYAS TIWARI & EUGENIA WILSON

PREPARED, SUPERVISED, INSTRUCTED AND REVISED BY: CHIH-LING TSAI

Table of Contents

| | |
|--|----|
| Preface | 36 |
| Data Resources | 36 |
| i. Individual Data Resources..... | 36 |
| ii. Data Repositories..... | 39 |
| iii. Data by Industry..... | 40 |
| iv. Book Resources..... | 42 |
| v. Artificial Intelligence / Machine Learning Data Resources | 42 |
| Conclusion..... | 43 |

Preface

When pursuing a project, it can be difficult to assess what sort of data you will need and where you will be able to access it. In this report, as students of the GSM, we aim to make these steps more straightforward for our peers, alumni, and anyone else who comes across this paper. The data files in this report are broken up into five sections: individual data resources, data repositories, data by industry, book resources, and Artificial Intelligence / Machine Learning. “Data by Industry” will encompass the prior resources by further elaborating on ways in which they can be utilized in the business environment. It is our goal to help the reader solve business-related problems by providing possible data resources in an easy to digest, straight-forward fashion.

Data Resources

8.1 Individual Data Resources

In this section, we have identified potential resources for data collection and elaborated on how to use the data within the resource. In the first column on the left-hand side, is the name of the data resource as well as a hyperlink to access it. While the majority of the data is free and easy to access, on some links it is also noted if the data has an associated cost or if the user will need to create an account / request access to the data. The second column presents an overview of the subject of the data, ranging from the specific region of the data (U.S. vs World) to descriptors such as economics, crime, and health. The third column describes the type of data, which will become more specific, such as monthly, quarterly, etc., as the data becomes more niche. The fourth and final column focuses on data utilization. While the majority of the resources contain time series data, additional ways of manipulating the data such as clustering, t-tests, and ANOVA are mentioned. It is important to note that the descriptors here are not all encompassing, and that some data subjects and types of data may be included in the links but not in the report.

| Data Source | Subject | Type of Data | Data Utilization |
|--|--|------------------------------|--|
| AWS Open Data https://registry.opendata.aws/ | | | Multivariate, univariate, sequential, time series, text, domain-theory; classification, regression, clustering |
| Bureau of Economic Analysis https://www.bea.gov/ | U.S. and World - GDP, Personal Income, International Trade | 1929-2020: yearly, quarterly | Time series data, regression |
| Congressional Budget Office https://www.cbo.gov/data/budget-economic-data | U.S. budget projects and data | 10-year budget projection | Time series data |
| Crypto Data https://www.cryptodatadownload.com/data/ | U.S. and World - Crypto exchanges and historical performance, machine learning, AI | Time series | Time series data |
| Data USA https://datausa.io/ | U.S. – city and congressional district, transportation, sports, retail, education, | Numerical, categorical | Multivariate, univariate, sequential, time series; classification, regression, clustering |

| | | | |
|--|---|--|--|
| | construction, Covid-19, machine learning | | |
| DatHub https://datahub.io/collections | U.S. and World - climate change, demographics, economic, GeoJSON, machine learning | Numerical, categorical | Multivariate, univariate, sequential, time series data; classification, regression, clustering |
| Directory of Radiotherapy Centers (IAEA) www.dirac.iaea.org/ <i>*data accessible on request for academic research, government, and international organizations only</i> | Data on radiotherapy and light-ion centers, teletherapy and brachytherapy machines, treatment planning systems and computed tomography units and simulators | Current and historical | |
| EU Open Data Portal https://data.europa.eu/en | European (36 countries) - economics, employment, science, environment, education | Numerical, regional | Time series data, multiple linear regression, one-way or two-way ANOVA |
| Federal Bureau of Investigation https://www.fbi.gov/services/cjis/ucr/ | U.S. – crime / hate crime data | Quarterly, yearly, geographic | Time series data, multiple linear regression, one-way or two-way ANOVA |
| Federal Reserve Bank of St. Louis https://research.stlouisfed.org | U.S. – economic indicators | Monthly, quarterly | Time series data |
| Federal Reserve Bank of San Francisco https://www.frbsf.org/economic-research/ | U.S. – economic indicators | Monthly, quarterly | Time series data |
| Google Trends https://trends.google.com/trends | Data by country - what type of traction (how many google searches) is a topic getting | Daily, weekly, monthly - since 2004 | Time series data |
| Health and Retirement Study https://hrs.isr.umich.edu/data-products <i>*must create account to access</i> | U.S - health and aging data (income, health insurance, pension plans, etc.) | Biennial (every 2 years), off-year, and cross-year data products; since 1960 | Time series data |
| Infoplease www.infoplease.com | U.S. and World - various (economic, population, birth rates, education, labor, etc.) | Ranking (top 10, 25, etc.), yearly | Time series data, multiple linear regression, one-way or two-way ANOVA |
| Inter-university Consortium for Political and Social Research www.icpsr.umich.edu/icpsrweb/index.jsp | U.S. and World - data covers a wide range of topics, Covid-19 | Much of the data is up to 2019 | |

| | | | |
|--|---|--|---|
| <i>*must be a member to access</i> | | | |
| Internal Revenue Service https://www.irs.gov/statistics | U.S. – tax collection data | Yearly | Time series data |
| International Monetary Fund https://www.imf.org/en/data | Global - fiscal data (debt, commodities, government, etc.) | Monthly, quarterly, regional | Time series data |
| Kaggle Datasets https://www.kaggle.com/datasets | Thousands of data sets on various topics (>50k datasets, >400k public notebooks) | Numerical, categorical | Time Series Data, multiple linear regression, one-way or two-way way ANOVA |
| NASDAQ Data Registry https://data.nasdaq.com/ https://www.nasdaq.com/market-activity <i>*market data is free (2nd link) but other data is not free</i> | Stock data, buy and sell orders in NASDAQ for securities listed on NASDAQ, NYSE, NYSE American | Monthly, yearly data up to 10 years | Time series data |
| National Bureau of Economic Research https://www.nber.org/ | U.S. – business cycle expansions and contractions | Data from June 1857 – February 2020 | Time series data |
| National Center for Health Statistics (CDC) www.cdc.gov/nchs/ | U.S. and U.S. Territories - birth and death rate data | Yearly | Time series data |
| Social Progress Index www.socialprogress.org | California, U.S., and World - social progress (shelter, personal freedom and choice, health and wellness, etc.) | Differences across locations for a year; can find data points for previous years | Time Series Data, multiple linear regression, one-way or two-way way ANOVA |
| Statista https://www.statista.com/ <i>*cost associated</i> | U.S. and World - various (technology, CPG, oil, travel, etc.) | Yearly, quarterly, rankings | Multiple linear regression, single variable for large samples, paired data for t-tests, data for one-way or two-way ANOVA, time series data |
| United States Census Bureau https://data.census.gov/cedsci | U.S. - various (economic, population, birth rates, education, labor, etc.) | Yearly data often 2014-2019; some data is only for a specific year | Time series data |
| United States Bureau of Labor Statistics https://www.bls.gov/ | U.S. – Labor and economic statistics | Monthly, yearly | Time series data |
| World Bank Data Registry https://data.worldbank.org/ | Global – development data | | Time series data |

8.2 Data Repositories

While we hope that the table above has provided you with a strong start, sometimes you just need to do more digging. In this section we have identified data repositories and the topic/s they cover, to hopefully guide you in the right direction. These resources range from 70 to 500+ additional data sources to help find what you are looking for.

| Data Repository | Summary |
|---|--|
| “Best Data Sources for Business Analysis” https://www.analyticsinsight.net/best-data-sources-for-business-analysis/ | Databases which cover topics like commercial data, time series data, news, legal, media, education, fun, government |
| Carnegie Mellon University: Machine Learning and AI https://guides.library.cmu.edu/machine-learning/datasets | Data repositories that cover topics such as ML, text mining, and COVID-19 |
| Columbia University Big Data Analytics https://www.ee.columbia.edu/~cylin/course/bigdata/getdatasetinfo.html | Large public datasets covering various sectors such as retail, media, finance, life science, government |
| Dartmouth Library https://researchguides.dartmouth.edu/datasets | Time series / data sets on financial data, social science, government |
| Forbes https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/?sh=67b03146b54d | Data resources, many of which are affiliated with U.S. and international government bodies |
| Free Data Resource Database https://guides.emich.edu/data/free-data | Database of websites that are data repositories |
| Google Dataset Search https://datasetsearch.research.google.com/ | Links to search-relevant data |
| <i>*costs may be associated</i> | |
| KD Nuggets https://www.kdnuggets.com/2017/12/big-data-free-sources.html | 70+ free datasets: AI, machine learning, R, Python, NLP, deep learning, data visualization |
| Nature https://www.nature.com/sdata/policies/repositories | Scientific data repositories covering biology, health, chemistry, physics, social sciences, climate |
| Registry of Research Data Repositories https://www.re3data.org/ | A global registry of research data repositories from all academic disciplines |
| UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php | 585 ML focused data sets - life sciences, physical sciences, engineering, social sciences, business, game |
| WorldData.AI https://worlddata.ai/ | 3.3 billion datasets across macroeconomics, trade, labor statistics, financial markets, weather, health and demographics |
| <i>*free for academics</i> | |

8.3 Data by Industry

In a further effort to streamline your search for the perfect data, this section was created. The data below has been categorized into what can be considered business functions but also industries, for when the term business unit will not suffice. The data in this section was extrapolated from the pertinent links above and distributed in the following format: Finance, Technology, Operations, Food and Agriculture, Social, and Government. Similar to the data repository section, each category is divided into two columns: Data Source and Summary, in the effort to provide a concise understanding of where to access the data and what kind of data can be expected.

| Data Source | Summary |
|--|---|
| Finance | |
| Bureau of Economic Analysis https://www.bea.gov/ | GDP, Personal Income, International Trade datasets |
| Congressional Budget Office https://www.cbo.gov/data/budget-economic-data | Datasets from hundreds of U.S. budget projects and datasets |
| Federal Reserve Bank of St. Louis https://research.stlouisfed.org | Large public datasets from Federal Reserve Bank of St. Louis |
| International Monetary Fund https://www.imf.org/en/data | Global - fiscal data about debt, commodities, government, etc. |
| Internal Revenue Service https://www.irs.gov/statistics | Time-series data from Internal revenue service records. |
| Crypto Data https://www.cryptodatadownload.com/data/ | Data on cryptocurrency exchanges and performance |
| Federal Reserve Bank of San Francisco https://www.frbsf.org/economic-research/ | U.S. – economic indicators |
| Technology | |
| Kaggle Datasets https://www.kaggle.com/datasets | Kaggle provides data from multiple industries. Thousands of data sets on various topics (>50k datasets, >400k public notebooks) |
| Google Dataset Search https://datasetsearch.research.google.com/ | Thousands of ML and user-based datasets from Google and related repositories. |
| AWS Open Data https://registry.opendata.aws/ | Thousands of ML and user-based datasets from AWS and related repositories. |
| UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php | 585 ML focused data sets - life sciences, physical sciences, engineering, social sciences, business, game |
| Google Trends https://trends.google.com/trends | Data by country - what type of traction (how many google searches) is a topic getting |
| Directory of Radiotherapy Centers (IAEA) www.dirac.iaea.org/ | Data on radiotherapy and light-ion centers, teletherapy and brachytherapy machines, treatment planning systems and computed tomography units and simulators |

| | |
|--|---|
| *Data accessible on request for academic research, government, and international organizations only | |
| Operations | |
| Kaggle Datasets https://www.kaggle.com/datasets | Kaggle provides data from multiple industries. Thousands of data sets on various topics (>50k datasets, >400k public notebooks) |
| National Bureau of Economic Research https://www.nber.org/ | U.S. – business cycle expansions and contractions |
| Food and Agriculture | |
| Kaggle Datasets https://www.kaggle.com/datasets | Kaggle provides data from multiple industries. Thousands of data sets on various topics (>50k datasets, >400k public notebooks) |
| USDA Datasets https://www.usda.gov/content/usda-open-data-catalog | 100+ Dataset from United States Department of Agriculture |
| Food and Agriculture organization of the United nations http://www.fao.org/faostat/en/#home | 100+ Dataset from Food and Agriculture organization of the United nations |
| Social | |
| Health and Retirement Study https://hrs.isr.umich.edu/data-products | (Login Required) 100+ datasets from U.S - health and aging data (income, health insurance, pension plans, etc.) |
| Datahub https://datahub.io/collections | U.S. and World - climate change, demographics, economic, GeoJSON, machine learning |
| National Center for Health Statistics (CDC) www.cdc.gov/nchs/ | U.S. and U.S. Territories - birth and death rate data |
| Social Progress Index www.socialprogress.org | California, U.S., and World - social progress (shelter, personal freedom and choice, health and wellness, etc.) |
| World Bank Data Registry https://data.worldbank.org/ | Global – development data |
| Social Progress Index www.socialprogress.org | California, U.S., and World - social progress (shelter, personal freedom and choice, health and wellness, etc.) |
| Government | |
| Federal Bureau of Investigation https://www.fbi.gov/services/cjis/ucr/ | U.S. – crime / hate crime data |
| USDA Datasets https://www.usda.gov/content/usda-open-data-catalog | 100+ Dataset from United States Department of Agriculture |
| International Monetary Fund https://www.imf.org/en/data | Global - fiscal data about debt, commodities, government, etc. |

| | |
|--|---|
| Congressional Budget Office https://www.cbo.gov/data/budget-economic-data | Datasets from hundreds of U.S. budget projects and datasets |
| Internal Revenue Service https://www.irs.gov/statistics | Time-series data from Internal revenue service records. |

8.4 Book Resources

The data resources below all stem from statistics textbooks. This section could prove useful not only when deciding what data to use, but also in regard to learning more about the best ways to utilize the data. Many of the data files in this section are also available in R.

| Book | Summary |
|--|---|
| <i>Applied Longitudinal Analysis</i> by Fitzmaurice, Laird & Ware (2004) http://www.hsph.harvard.edu/fitzmaur/ala/ | Datasets cover longitudinal studies on the development, persistence, and treatment of disease |
| <i>The Elements of Statistical Learning (2nd ed.)</i> by Hastie, Tibshirani & Friedman (2009) https://web.stanford.edu/~hastie/ElemStatLearn/ | Data predominantly covers bioinformatics and marketing - the book also comes with various R packages |
| <i>Longitudinal and Panel Data: Analysis and Application in the Social Sciences</i> by Frees (2006) https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Longitudinal%20and%20Panel%20Data/Book/DataFiles.htm | Data focuses on social sciences (e.g. insurance, tax fillings, divorce, donations, lottery) |
| <i>Semiparametric Regression</i> by Wand & Carroll (2013) http://matt-wand.utsacademics.info/webspr/data.html <i>*Some datasets are not publicly available</i> | Data covers topics like biological sciences, climate, retirement planning and trade union membership – data files are also available in R |

8.5 Artificial Intelligence / Machine Learning Data Resources

The data resources in this section primarily focus on the advanced Artificial Intelligence and Machine Learning datasets. These data resources can be used for various predictive analytical projects. Each of these data resources ranges from 10 to over 100,000 datasets in the similar domain. While analyzing these datasets, make sure that you take the date of data generation in account.

| Data Source | Summary |
|---|--|
| Allen Institute for AI, CORD-19 https://www.semanticscholar.org/cord19 | Over 44,000 scholarly articles, including over 29,000 with full text, about COVID-19 and the coronavirus family of viruses for use by the global research community. |
| Amazon Review Datasets https://snap.stanford.edu/data/web-Amazon.html | This dataset consists of reviews from amazon. The data span a period of 18 years, including ~35 million reviews up to March 2013. Reviews include product and user information, ratings, and a plaintext review. |

| | |
|--|--|
| Carnegie Mellon University: Machine Learning and AI https://guides.library.cmu.edu/machine-learning/datasets | Data repositories that cover topics such as ML, text mining, and COVID-19 |
| Cityscape Datasets https://www.cityscapes-dataset.com/ | This is an extensive dataset that has street scenes in 50 different cities. The Cityscapes Dataset focuses on semantic understanding of urban street scenes. In the following, we give an overview on the design choices that were made to target the dataset's focus. |
| Google Open Image Dataset https://ai.googleblog.com/2016/09/introducing-open-images-dataset.html | Open Images, a dataset consisting of ~9 million URLs to images that have been annotated with labels spanning over 6000 categories. |
| Nature https://www.nature.com/sdata/policies/repositories | Scientific data repositories covering biology, health, chemistry, physics, social sciences, climate |
| Open Data for Deep Learning https://wiki.pathmind.com/ | Maintained by a model deployment platform, Skymind. Has a collections of open datasets. |
| UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php | 585 ML focused data sets - life sciences, physical sciences, engineering, social sciences, business, game |
| Waymo Open dataset https://waymo.com/open/ | This is a dataset resource from the folks at Waymo. Includes a vast dataset of autonomous driving, enough to train deep nets from zero. |

Conclusion

It is our sincere hope that the above data resources and various categorizations, have streamlined your search for the most relevant and useful data. Data is an invaluable tool to making key decisions and further understanding the inner workings of the world around us. These datasets were identified to not only be utilized by GSM students for individual or classroom projects, but to continually support alumni in their efforts to solve challenging problems and provide clarity on complex situations. This Data Resources Handbook was designed to be a launching pad, and as data collection and analysis evolves, we hope this resource evolves with it.