SUITESPOT

**To:** All Executives and Hotel Managers

**From:** Decision Support Systems

**Subject:** Data Mining Recommendations

**Date:** October 8, 2017

---

# 1 Overview

## 1.1 What is data mining?

Data mining is the process of identifying information from vast amounts of data. Data mining utilizes employees and automated programs to analyze an organization's data looking for trends, patterns, anomalies, or outliers, with the hope of identifying actionable insights from a set of data. Data mining seeks to transform the data that a business collects into information that can used in ways such as better understanding their customers, evaluating product sales performance, or even improving processes. Data mining harnesses practices from statistics, artificial intelligence, data visualization, and business analysis, all of which contribute to the continued growth in popularity of data mining. Those who conduct data mining will need at least a high-level understanding of statistics, a familiarity with computers and programming, and an adequate amount of professional skepticism to identify hidden insights in the data. A key element of data mining is asking good questions and answering them using the available data, a topic we will cover in further detail soon with an example listing of questions included.

Regarding data mining, Thomas Davenport in the Harvard Business Review details, "The latest strategic weapon for companies is analytical decision making." [1] Davenport further explain companies "have used analytics to better understand their customers and optimize their extended supply chains to maximize their returns on investment while providing the best customer service." [1] Most importantly, Davenport concludes "the level of success is highly dependent on a company understand its customers, vendors, business processes, and the extended supply chain," [1] which can be done using data mining.

## 1.2 What is an example of data mining?

To highlight the value of data mining, we will next demonstrate the potential value of data mining with a couple key examples. The first example demonstrates the value of data mining at an industry leading retailer and the second at an industry leading hospitality chain. In these examples, we will consider how each company has used data mining to drive overall business growth.

A great, easy to comprehend example of data mining is data mining used in retail stores on customer and transaction data. Upon analyzing this data, data mining can reveal unknown insights regarding customer shopping habits, transaction size or frequency, and even specific product sales details. A excellent real-world application of data mining comes from Target. As described by Forbes "Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy." [2] One of the ways Target uses data mining is to try "to hook parents-to-be at that

crucial moment before they turn into rampant - and loyal - buyers of all things pastel, plastic, and miniature." [2]

As such, Target applies data mining techniques to their customer details and buying habits to identify customers on the verge of becoming parents, sometimes before the customers even realize it themselves. Target created a 'pregnancy score' based on customer buying habits of key items such as unscented lotion, supplements, and big bags of cotton balls. Based on the pregnancy score, Target began sending automatic mail coupons for nursery furniture and maternity clothing to customers exceeding a certain score. The article further sites an instance where Target new one man's high school age daughter was pregnant before he did. The man found out about the pregnancy after seeing Target baby clothes coupons in the mail. Target used data mining to better understand their customers and in this case, because of data mining, Target advantageously knows more about their customers than their customers know about themselves.

A more specific example of data mining in the hospitality industry comes from the US hotel chain Red Roof Inn. "During the record-setting winter of 2013/2014, [Red Roof Inn] realized the huge value of having a number of hotels close to major airports at a time when flight cancellation rate was around 3%. This meant around 90,000 passengers were being left stranded every day. The chain's marketing and analytics team worked together to identify openly available public datasets on weather conditions and flight cancellations. Knowing that most of their customers would use web search on mobile devices to search for nearby accommodation, a targeted marketing campaign was launched, aimed at mobile device users in the geographical areas most likely to be affected. This led to a 10% increase in business in areas where the strategy was deployed." [3] Red Roof Inn employees utilized existing, publicly available data to analyze weather conditions and flight cancellations to issue targeted advertisements leading to a significant increase in business. This example is data mining at its best.

### 1.3 What would be the benefits of data mining to the company?

The chief benefit of data mining is identifying actionable information that is advantageous to the organization that may not have been discovered otherwise. Companies collect troves of data through a myriad of sensor, device, and input network each day. Unfortunately, much of that data is never used. Data mining puts this previously untapped data to use by uncovering hidden value, unknown insights, and quite unexpected realizations within the data. In describing the benefits of data mining, Dr. Arno Penzias explained, "If you're not doing this, you're out of business." [1] Dr. Penzias frames data mining not exactly as a benefit, but more so a necessary means of survival in today's hypercompetitive business marketplace.

### 1.4 What questions might the company attempt to answer using data mining?

Data mining is often hypothesis-driven, which allows the company to ask and seek answers a plethora of critical questions. As previously mentioned, data mining is often best used when employees first ask questions, and then answer them with the data. By using hypothesis testing, companies can use the data they already have to answer the outstanding questions necessary to improve their business. Asking good questions are key to hypothesis-driven data mining. Included are several questions SuiteSpot may attempt to answer using data mining:

1. What are the most popular dates for dinner at each hotel location?
2. As storing inventory, particularly perishable items, for an extended period of time can be costly, what appetizers, entrees, and desserts are the least popular?
3. What desserts are most often selected with what entrees?

4. What appetizers and entrees sell the most wine along with the dinner?
5. How many tables does each restaurant turn over in a night and how fast on average does each table turn over?
6. What is the average table size in each restaurant and what is the average party size at each restaurant?
7. Do certain menu items, either food or drinks, sell better in certain locations?
8. What is the most common payment method used and does payment method have any relation to tip amount?
9. Does the day of the week of the dinner date correlate to any food or drink purchases?
10. Which restaurants receive the highest amount in tips and which food and drink items correlate with the highest amount in tips?

### 1.5 Finally, what information technology infrastructure needs to be in place before the organization can actually perform any data mining activity?

To even perform data mining, SuiteSpot will first need to have adequate information technology infrastructure on hand. On the software side, SuiteSpot will need the appropriate data mining software. Many big players in the software world publish their own data mining software such as IBM SPSS, Statistica, and SAS. SuiteSpot would also be able to use business intelligence software such as SAP, IBM Cognos, or Teradata for data mining although their capabilities are limited when compared to data mining software. SuiteSpot will also likely want to have a software platform and analytics code language such as R, SQL, or Python so employees can conduct their own, ad-hoc type analyses. Additionally, SuiteSpot will need to have their data stored in a data warehouse or a collection of data marts so these software tools will be able to analyze the data.

On the hardware side, SuiteSpot will need server space to store the data and hold the software necessary for data mining. SuiteSpot has plenty of options in this arena and they can choose from a physical/on-premise hardware architecture, a cloud-based architecture such as Amazon Web Services (AWS), or a mix of both. SuiteSpot should decide on their hardware strategy first as without any hardware to back it up, software is useless.

Beyond software and hardware, SuiteSpot will need adequately skilled employees to conduct the data mining. The information technology focus here is people, processes, and technology. Without the right people, the process of data mining and the hardware and software technology is meaningless. As previously mentioned, data mining expertise draws from the fields of statistics, programming, data analytics, and even business. To make the information technology infrastructure most effective, SuiteSpot will need to higher individuals with experience as statisticians, computer scientists, data analysts, data scientists, or business analysts. To summarize, the data mining solution for SuiteSpot involves three key pieces: people, processes, and technology, with each piece being necessary for success.

## 2 Data Mining

We now present a demonstration of the value of data mining on data provided by Suite Spot consisting of dining transactions at five Suite Spot locations. Often we will not be given a particular task with the data, but we will be asked to find something interesting or useful about it. It is crucial to first understand the data. First we may want to look at the size of both the dimensions and the observations. This dataset contains 12 dimensions and 1011 observations.

We must first understand the dataset from a structural standpoint. We can quickly do this by looking at the first few observations.

| Date | Location | Table Size | Appetizer | Salad | Entree | Dessert |
|---|---|---|---|---|---|---|
| 03/25/2017 | 1 | 2 | Lettuce Wraps | Vinaigrette | Vegetarian | Cheesecake |
| 09/04/2016 | 3 | 3 | None | House | Chicken | Cheesecake |
| 08/19/2017 | 5 | 6 | Lettuce Wraps | 1000 Islands | Chicken | Fruit |
| 09/11/2016 | 4 | 2 | None | None | Vegetarian | None |
| 02/07/2017 | 4 | 6 | Lettuce Wraps | None | Chicken | Fruit |
| 06/03/2017 | 3 | 4 | Buffalo Wings | House | Steak | Fruit |

continued...

| N.A. Drinks | Wine | Other Drinks | Payment Method | Tip |
|---|---|---|---|---|
| Water | Chardonnay | Rum | Credit Card | Below 10% |
| Fruit Juice | Shiraz | Gin | Credit Card | 15-20% |
| Pellegrino | None | None | Credit Card | 15-20% |
| Pellegrino | Chardonnay | None | Credit Card | 10-15% |
| Water | Chardonnay | Rum | Credit Card | 15-20% |
| Diet Soda | Merlot | Rum | Cash | 15-20% |

Here, we note the entries have a date, a location ID, the size of the table (number of patrons), the various meal and drink items the patrons chose, the payment method, and the tip bucket. After performing some data cleaning, and removing some observations with missing dates, we can then group the observations into buckets by month. This allows us to aggregate and visualize the data. By visualizing our dataset we are able to quickly determine where to focus our efforts.

## 2.1 Exploratory Data Analysis

There are a lot of ways to visualize data. We may wish to start with simple frequency plots to see what food and drink items stand out. For drinks, we omit occurrences where water or no drinks were ordered, since we are only interested in purchases. There may be situations however, where we are interested in when water or no drinks were ordered, as the absence of a purchase has different implications. For now, we will focus on purchases. Figure 1 shows steak to be our bestselling entree, and cheesecake stands out as our bestselling dessert. Conversely, apple pie, is not selling well as a dessert. It may be a good idea to look into the reasons behind this. Our worstselling item overall appears to be the crab cakes appetizer. We may need to look at either improving the quality of the item, or replacing it altogether.

From the perspective of drinks, in Figure 2, margarita sales appear to be a little low, but chardonnay dominates our overall sales. It may be possible to increase the price on chardonnay while offering happy hour specials on margaritas, and in doing so, increase profits.

### 2.1.1 Trends Over Time

Looking at our overall patronage in Figure 3, there seems to be a dip in June, and a surge from October - December. We don't have enough data to determine the cause for the surge in the fall, but it would be worth looking into. Are there more events such as conventions, concerts, or festivals in the area around that time? This would be good to identify so we can determine our primary sales drivers. The dip in June is somewhat concerning. Can we improve our summer numbers by adding patio seating? This is worth investigating further.

Seasonality can often influence beverage choices so it may be interesting to see how beverage purchases behave over time. In Figure 4, it would appear that chardonnay sales tend to spike in the fall, but as we have seen, our overall patronage spikes in the fall, so a more valuable statistic to use might be the *rate* at whcih drinks sell. That is, the total number of a particular drink divided by the total number of patrons in a given month.

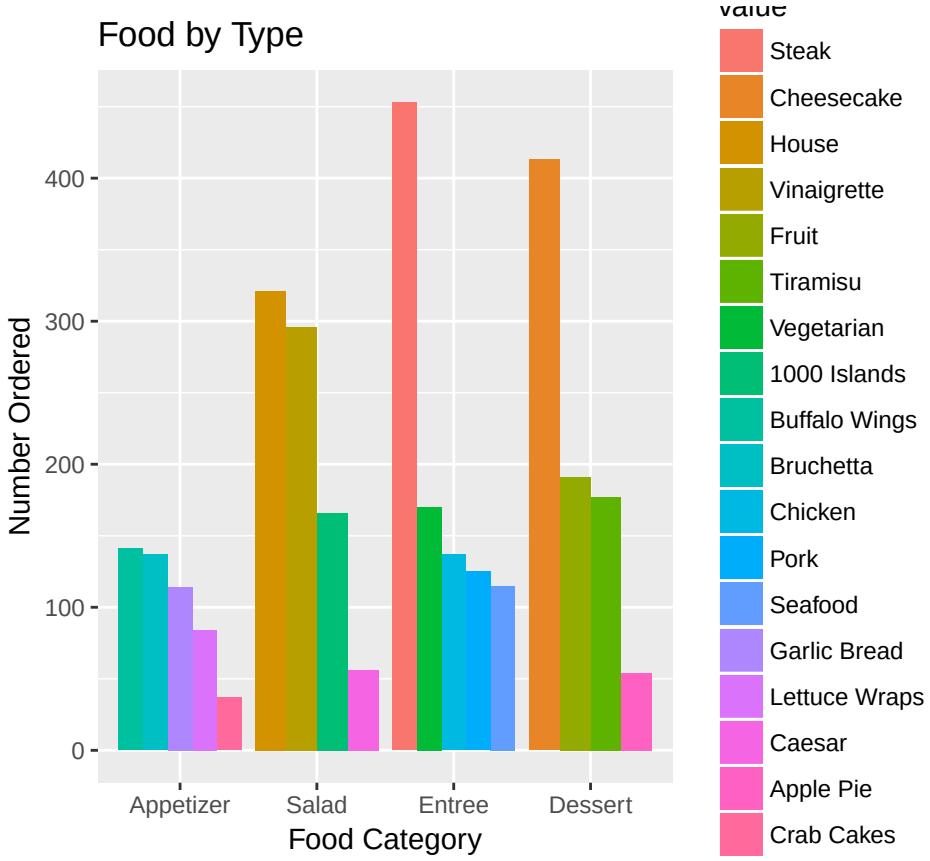


**Figure 1:** Food orders for all locations. Steak and cheesecake are standouts in their categories.

This is shown in Figure 5, and it is interesting to note that chardonnay rates actually peak in May. The overall chart in Figure 4 may be useful in managing inventory, but the rate may be more useful in identifying trends of popularity. If we wished to temporarily increase the price of chardonnay, May would be a good time to do so. This same technique can be applied to food as well.

We can also combine columns to determine the most popular item pairings. Our most popular item pairing is the **steak entree** and the **cheesecake dessert**. When steak was ordered, 29% of the time cheesecake was ordered as the dessert. Steak was also ordered with **chardonnay** 29% of the time, and a **house salad** was ordered with steak 24% of the time. We could use this information to offer a three course meal with wine as a special on Valentine's Day, for instance.

### 2.1.2 Evaluating Service

We can also discover *latent*, or hidden effects. If we wished to directly evaluate service quality at each of the locations, we would have to have patrons fill out surveys. This is not ideal, as we would have to account for non-response bias, and the logistics of encoding the responses. We may just not get very many responses at all, as customers generally just want to eat and drink, rather than fill out surveys. But we suspect that patrons tend to voice their opinions on service by the way they tip.

A good tip is generally considered to be at least 15% or more, and a bad one less than that. We can then simplify our tip buckets into "good" and "bad" tips. By examining the percentage of "good" tips at each location, we can determine which locations are doing well, and which ones need improvement. In Figure 6, we can see that our St. Louis location is well above the rest of the locations, and Memphis is lagging behind. We could theoretically use this data

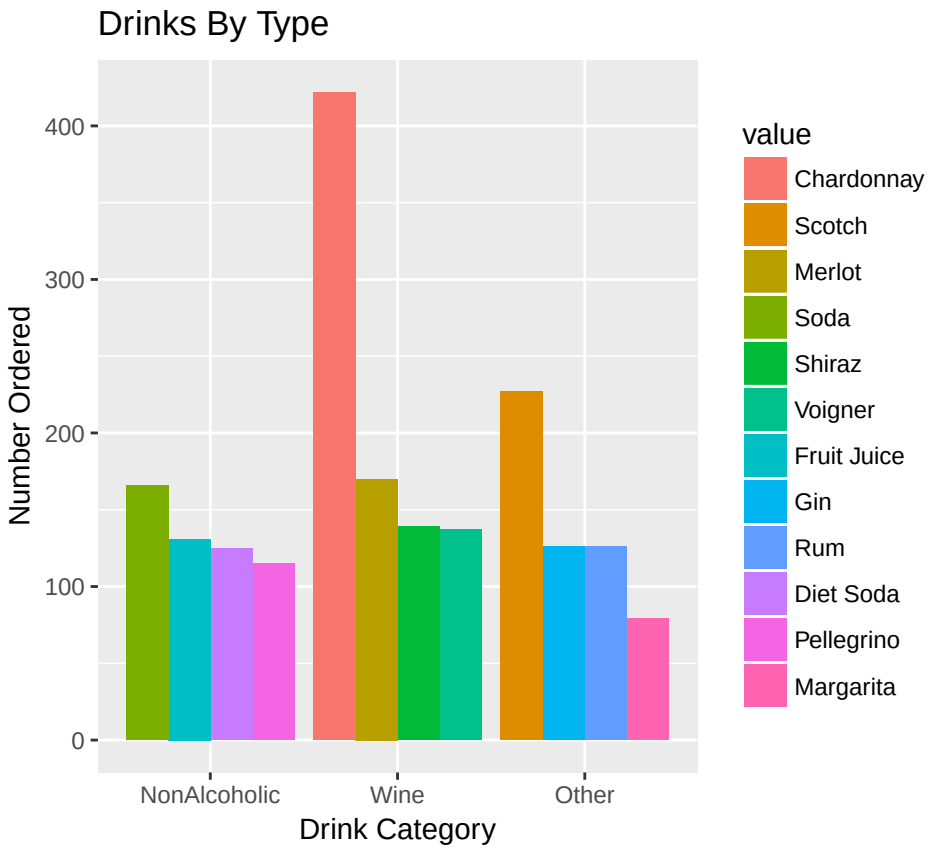


**Figure 2:** Drink orders for all locations. Chardonnay and scotch are standouts in their categories.

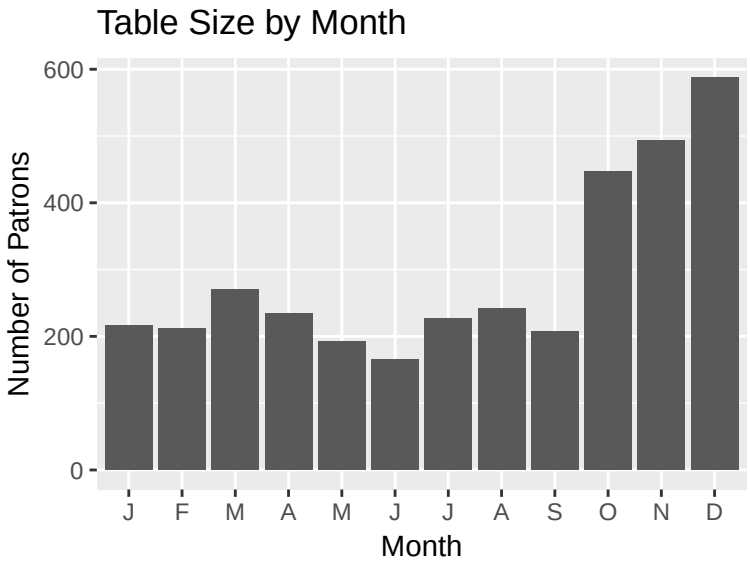


**Figure 3:** Overall patronage by month.

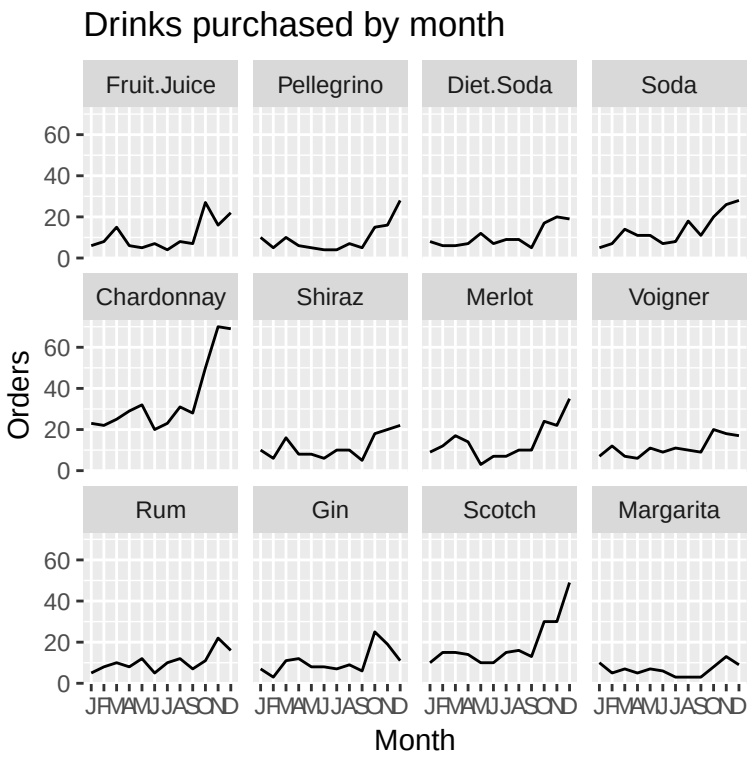


**Figure 4:** Overall drink purchases by month.

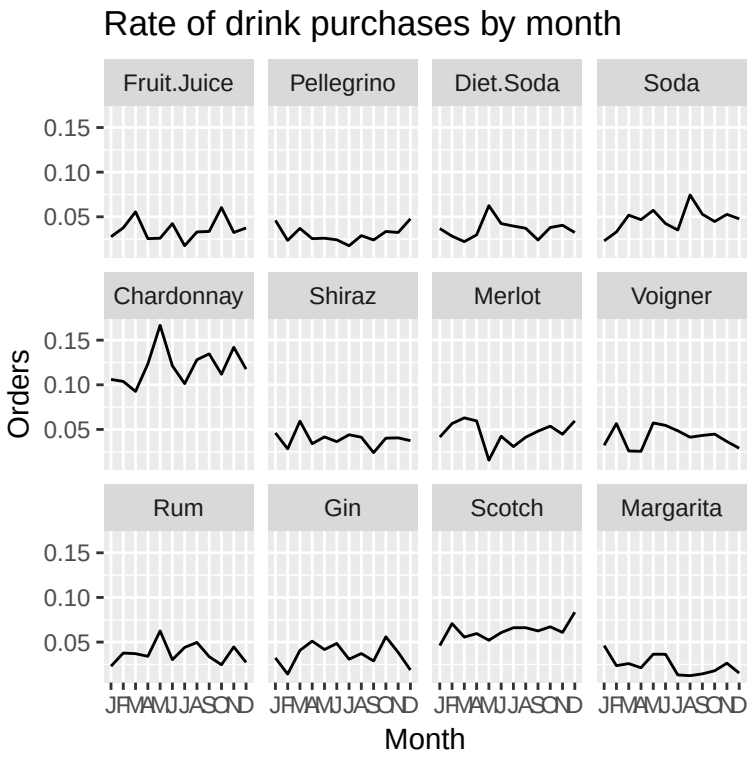


**Figure 5:** Drink purchase rate by month.

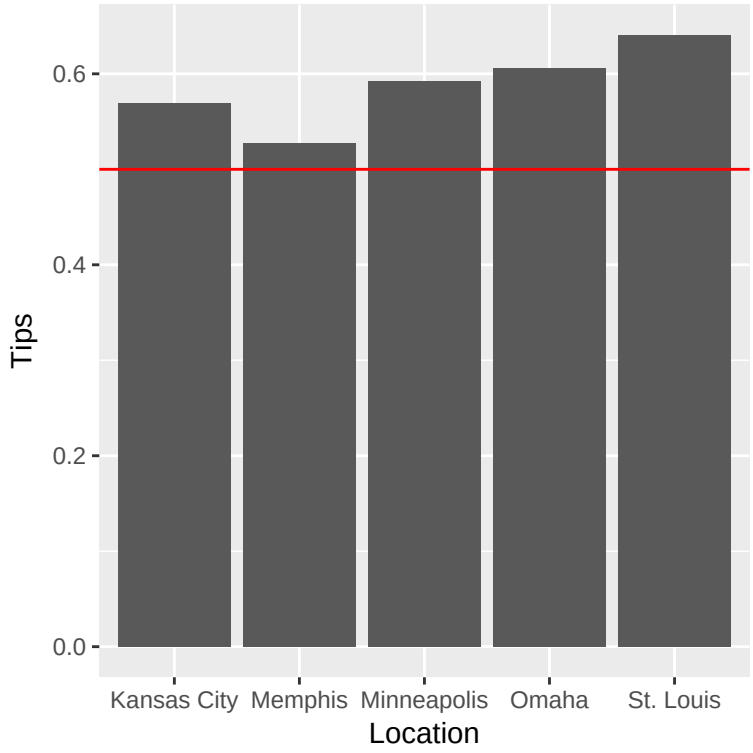


**Figure 6:** Good tips percentage by location by month. The red line indicates 0.50.

to drive incentive programs. Perhaps a "Location of the Month" contest, where we reward the location with the highest percentage of "good" tips could improve customer service. This would be a win-win as our wait staff would be receiving higher tips and our customers would be getting better service.

Perhaps those who purchase certain items are more likely to tip better than those who purchase others. While there did not seem to be enough evidence that indicate any of these item purchases could predict a good tip, some of the items do seem to fall above or below the mean proportion of good tips. We can see in Figure 7 patrons who purchase Buffalo Wings as their appetizer, are more likely to tip well, that is, 15% or above, while those who purchased rum as their drink, were slightly less likely to tip well, but over half of all patrons who purchased any of the items was more likely to tip well than not. Overall, the likelihood for a party being "good tippers" was about 59%, so we see this as a good news story.

## 3 Conclusion

While mining data through each of the location, steak stands out as a top seller for entrees but is closely followed by chicken as well as vegetarian options. A conclusion that can be drawn by this data is that steak is prepared well, and based on typical meat price points, customers are willing to spend money for a well-prepared entre. To improve the order rates of the other entrees, recipes can be improved and advertising can be increased. However, under the assumption that steak is the highest priced entre, each location is doing well with maintaining the highest profitability possible. In order to lower costs, each location could build partnerships with vendors providing this type of meat in hopes of receiving lower prices and turning a higher profit.

For Appetizers at each location, there are a lot of people not choosing an appetizer before their salad or entre. This can be concluded that appetizers are not well advertised on the

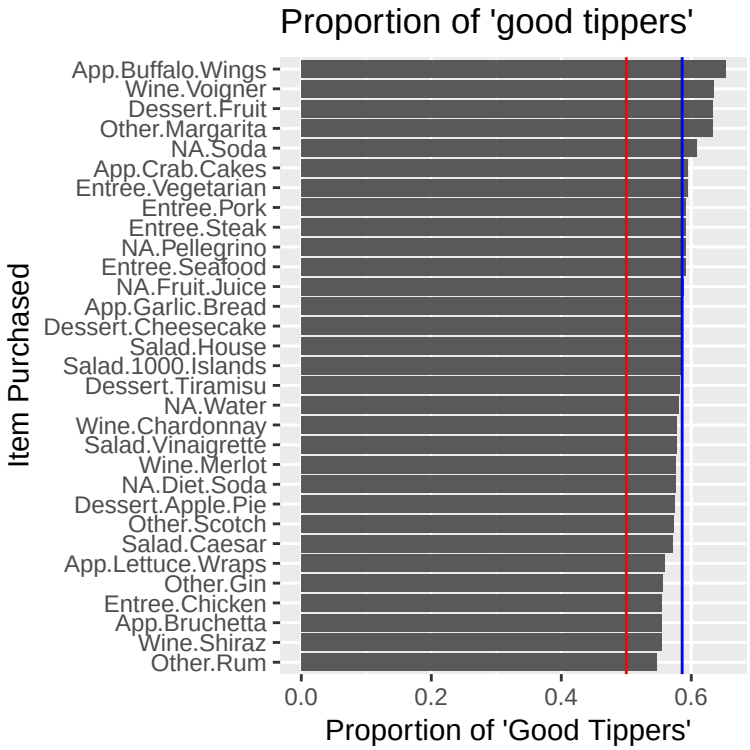


**Figure 7:** Good tips percentage by items purchase. The red line indicates 0.50, and the blue line indicates the mean.

menu or within a restaurant. Because the restaurant needs to be prepared for someone to order an appetizer, there is a risk that ingredients are being wasted - therefore lowering the potential profits of the restaurant. A conclusion that can be drawn from this data is to increase the pairings of the appetizers and entrees and encourage customers to purchase an appetizer through "paired" deals.

For desserts, cheesecake is the definite choice for the customers at each location across the board. The conclusion of this data can be drawn from the cheesecake well-prepared dish as well as perhaps "word of mouth" advertising. Most restaurants experience a "chain" reaction for customer orders. Once one person speaks about the quality of a dish, more people are willing to try it. Chardonnay is also a popular pairing at each location with the cheesecake. In order to increase profit, the restaurant could do a cheesecake and chardonnay pairing to increase customer dessert activity.

As stated before, there is not a specific correlation of data between each location. There are definitely top sellers and lower sellers, but each entre, desserts, and appetizer has a decent showing at each location. We believe advertising and entre and appetizer pairings are a good idea to increase the profit of each restaurant. Tipping is sufficient at each restaurant for the most part which shows that customer service is doing well. We conclude that the menu and restaurant appeal would help to increase profit and lower costs.

## References

[1] R. Sharda, D. Delen, and E. Turban. *Business Intelligence and Analytics*. Pearson, Upper Saddle River, NJ, 10 edition, 2015.

[2] Kashmir Hill. How target figured out a teen girl was pregnant before her father did. *Forbes*, 4 2017.

[3] Bernard Marr. How big data and analytics are changing hotels and the hospitality industry. *Forbes*, 1 2016.