Python for Business Analytics Data Analysis

By: Kasfa Mahi and Jesús Villegas
Professor John Droescher

Project Objectives:

The primary goal of our project is to assess and establish performance targets for the upcoming fiscal year. We aim to place a significant emphasis on aligning key metrics that will enhance the overall performance of the company, with a focus on improving and ensuring customers are more likely to return and repeat business. To achieve this, we have identified specific pain points from a customer's perspective, recognizing that addressing these issues will not only bring in customers continuously whether it be the same customers or new but also contribute to the overall success of the business.

Possible Pain Points Identified:

- **Late Deliveries:** Impacts the customers experience by leading to customer dissatisfaction and possible frustration. Additionally, late deliveries impact our reputation and how we look as a business. There may also be a specific problem, such as delivering outside of the continent or something such as a specific item being constantly out of stock.

By addressing these pain points, we anticipate not only enhancing the customer experience but also positively impacting the company's performance metrics. Improving the reliability of deliveries and ensuring adequate stock levels will contribute to customers being more likely to return, repeat business, and ultimately, the overall success of the company.

Our Process:

Reframe the question: Our original task was as follows, “I am your Regional Director and have presented you with the following question: I am looking to improve this year’s performance, what suggestions do you have for success?” Given we were tasked with providing suggestions on how to enhance this upcoming year’s performance. As a team we collectively thought of adapting a customer-centric approach, our reasoning being as follows, loyal customers are integral to a successful business. Taking that into account, we reframed the question as, "How can we strategically align key metrics for the fiscal year, emphasizing loyal customers, to evaluate performance targets, ensure progress tracking, and enable the company to make informed decisions and drive improvement in the targeted areas?"

Gathering Data: Our data originated from Professor John Droescher's database. Our primary emphasis was on collecting data related to late deliveries, customers, products, customer regions, customer states, and shipping methods.

Data Cleaning: Most of the data we used required no manipulation, as it was mostly clean. However, during the coding process, we did make some adjustments by renaming and removing specific titles to facilitate data manipulation.

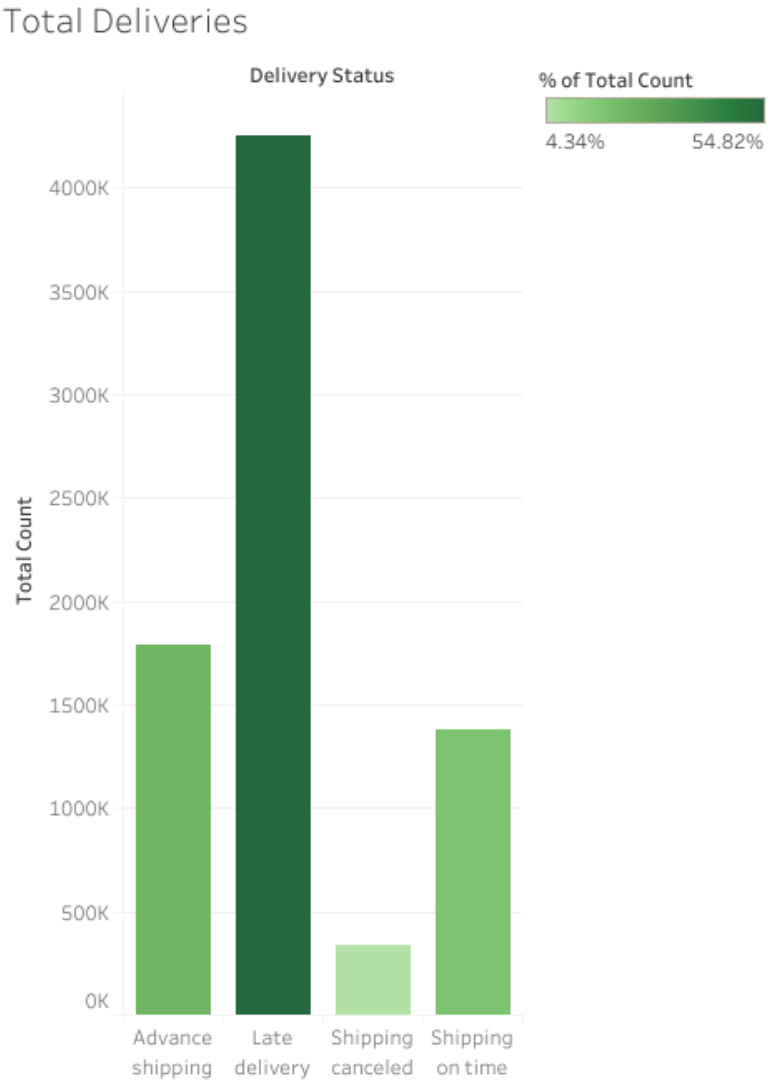
Data Visualization: In order to create visually compelling representation, we employed a dual approach, by harnessing both Excel and Tableau. Leveraging the versatile capabilities of Excel alongside the advanced visualization features of Tableau, we sought to design graphs that not only showcase relationships but also ensure a seamless and aesthetically pleasing interpretation of the data. This strategic combination allowed us to achieve a dynamic visual presentation, enhancing the accessibility and clarity of the story we aimed to convey.

Report: Our Final Report thoroughly details our task approach, encompassing data gathering, cleaning, and the data visualization process. Expanding on the analysis of identified trends, the report concludes with strategic future recommendations.

Presentation: We utilized our presentation in order to provide a snapshot analysis of trends and to summarize our findings as well as provide future recommendations.

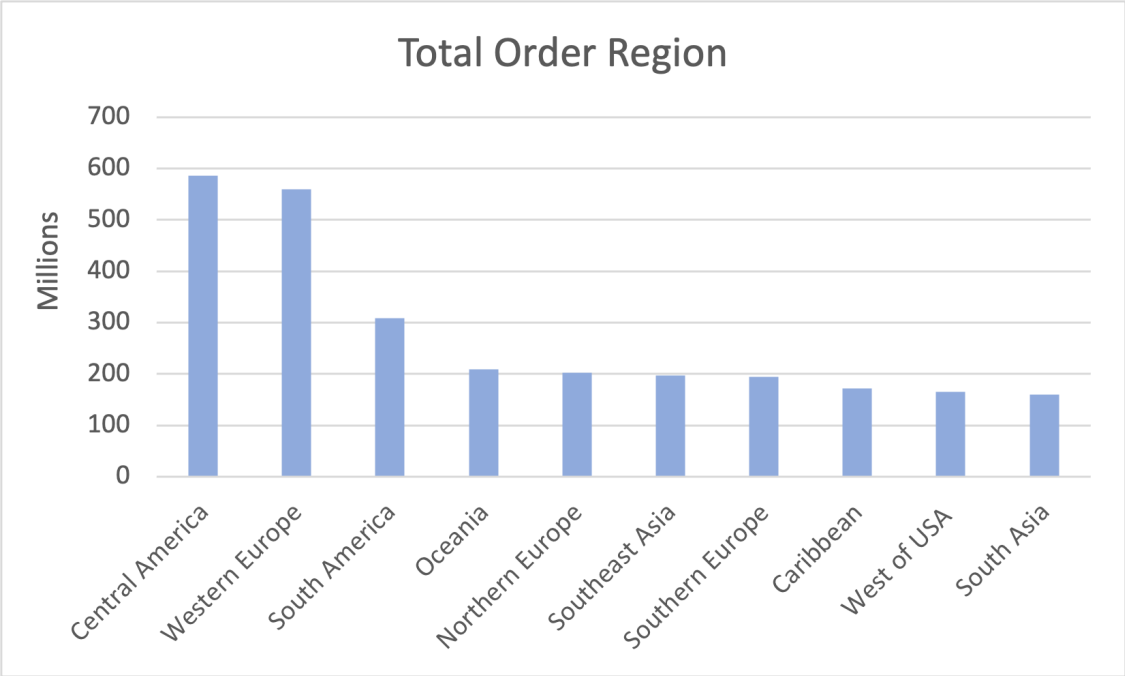
Report:

Task: Identify Potential Pain Points Related to Late Deliveries



Approach: In addressing the task, we formulated a query focusing on the "Delivery Status" and "Total Count" parameters. This query allowed us to identify the number of deliveries falling into distinct categories, including those shipped on time, canceled, and late. The data comprises information on delivery status, categorized into four segments: "Shipping on time," "Shipping canceled," "Advance shipping," and "Late delivery." The corresponding total counts for each category are as follows.

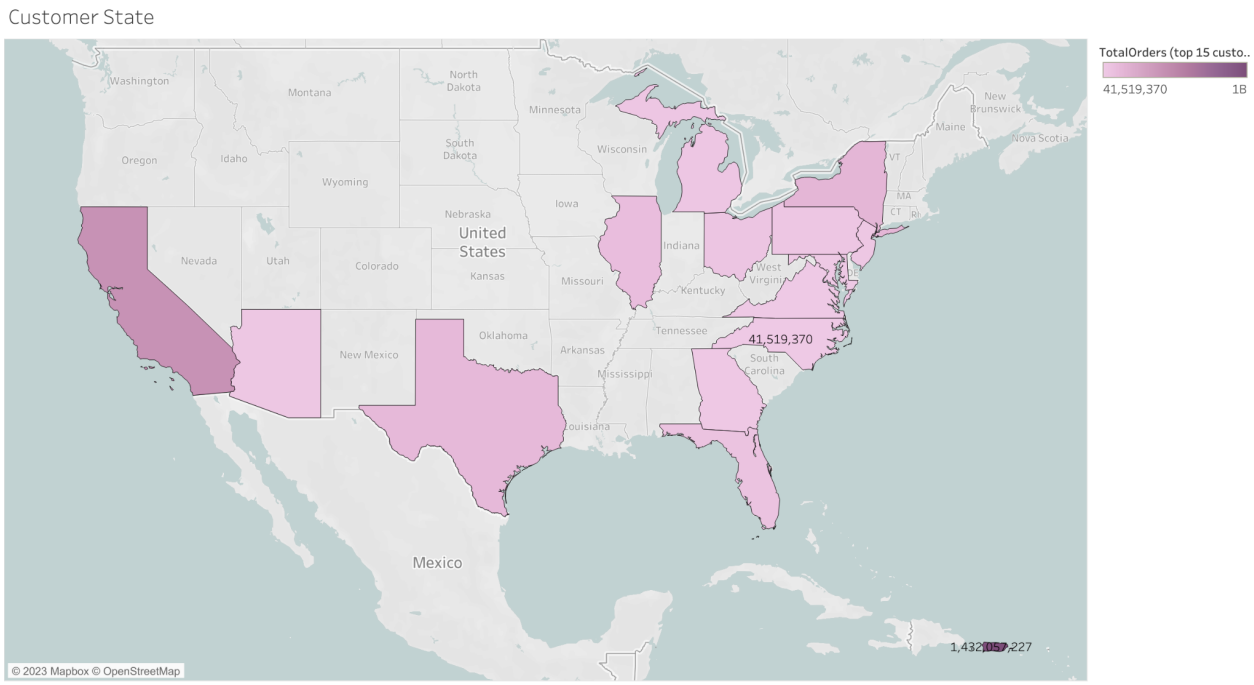
Data Overview: Our perspective began with the customer experience, revealing a significant pain point: a concerning frequency of late deliveries of 55% while only 18% of orders were being shipped out on time.



Task: Identify any correlation between Regional Dynamics and Late Deliveries

Approach: In response to recognizing the substantial volume of late deliveries, our task is to gain insights by visualizing the distribution of orders across regions. We aim to understand the dynamics of order distribution, specifically exploring how it may be influencing challenges related to late deliveries in various regions. Our approach involves formulating a query that emphasizes the "Order Region" and "Total Orders" parameters, allowing us to create a visual representation of the data. By analyzing this distribution, we can inform our strategy for addressing issues associated with late deliveries across different regions. The data, as seen above, will serve as a foundation for our exploration and subsequent decision-making.

Data Overview: The dataset provides information on total orders across different regions, showing a majority of the orders come from Central America and Western Europe.

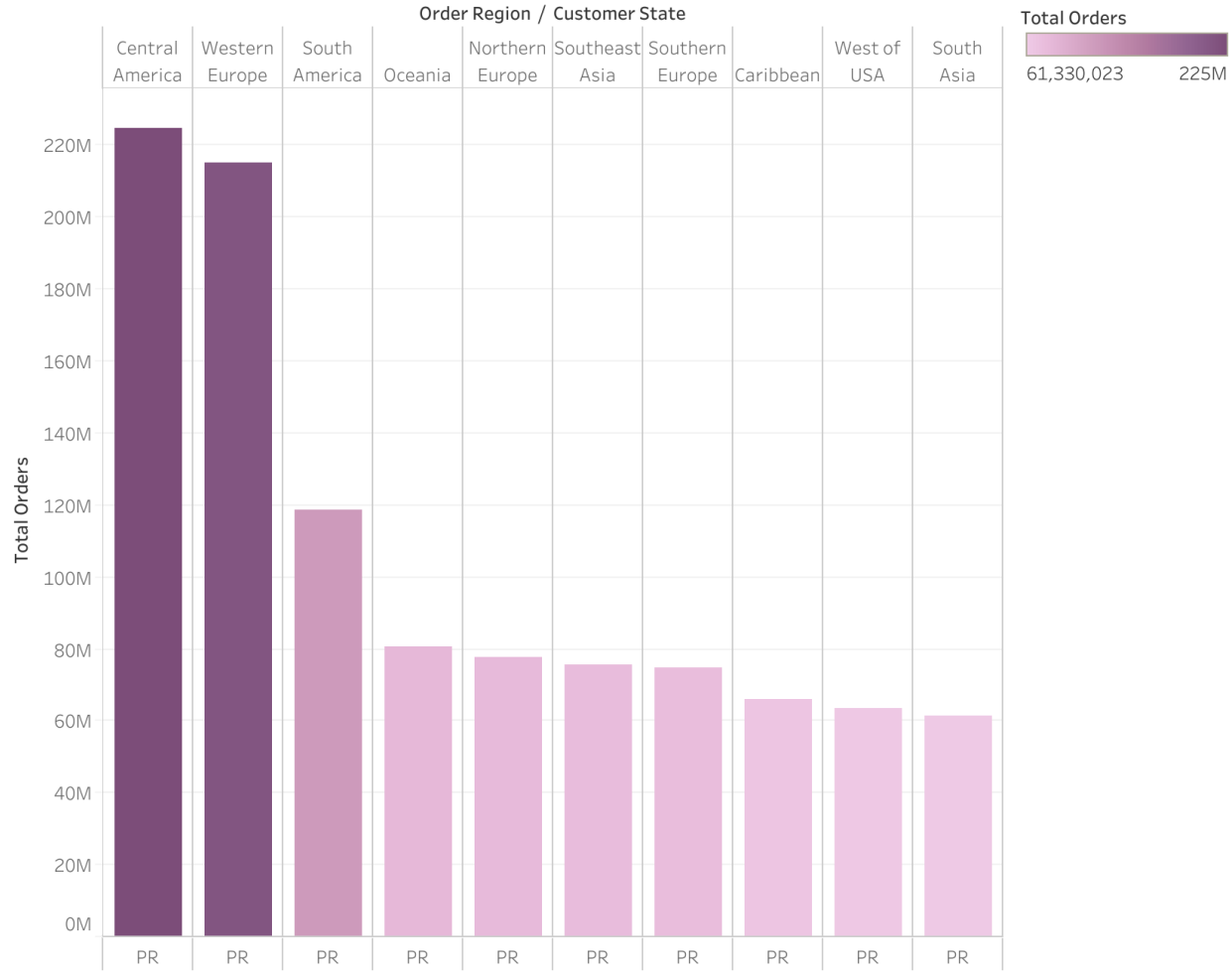


Task: Generate a visual representation to assess the impact of demographic factors on late deliveries, focusing on specific regions. For instance, examine if Puerto Rico (PR) experiences both a high volume of deliveries and a notable frequency of late deliveries, potentially due to challenges in shipping to that particular area .

Approach: To tackle this task, we investigated the potential influence of demographic factors on late deliveries. Recognizing a substantial occurrence of late deliveries, we first examined the shipping locations of all orders using the Total Order Region Graph presented on the previous page. In constructing a visual, our objective was to pinpoint which regions had the most orders per state. We utilized “Customer State” and “Order Id” as “Total Orders” to identify the top 15 states where total orders were being shipped.

Data Overview: The graph showed Puerto Rico with the highest number of late deliveries with 1,432,057,227 orders shipped to Puerto Rico and California following with 598,962,042 late orders.

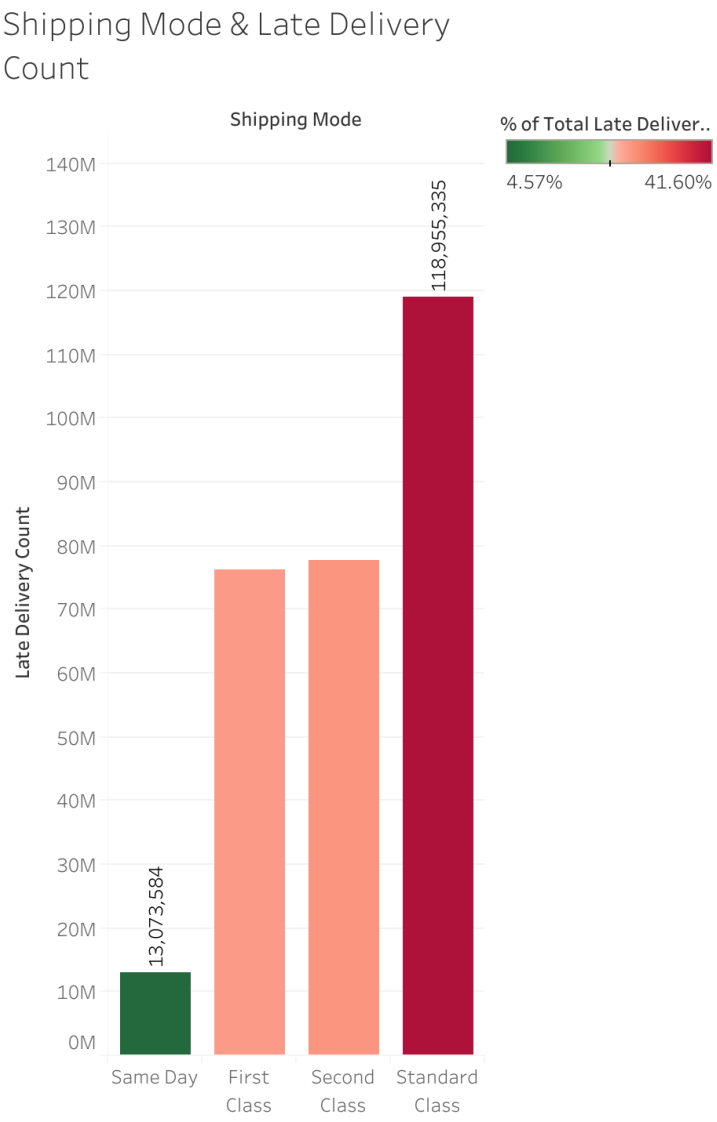
Order Region VS PR



Task: Our objective was to investigate now all the late delivery counts from the top 10 order region to Puerto Rico with the most shipped items.

Approach: The focus is not solely on highlighting the percentage of late deliveries by order region but rather on identifying patterns that may indicate potential challenges associated with overseas shipments, including flight delays and shipping issues. We employed “Order Region” and “Total Orders” to pinpoint the top 10 regions responsible for shipping late deliveries to Puerto Rico.

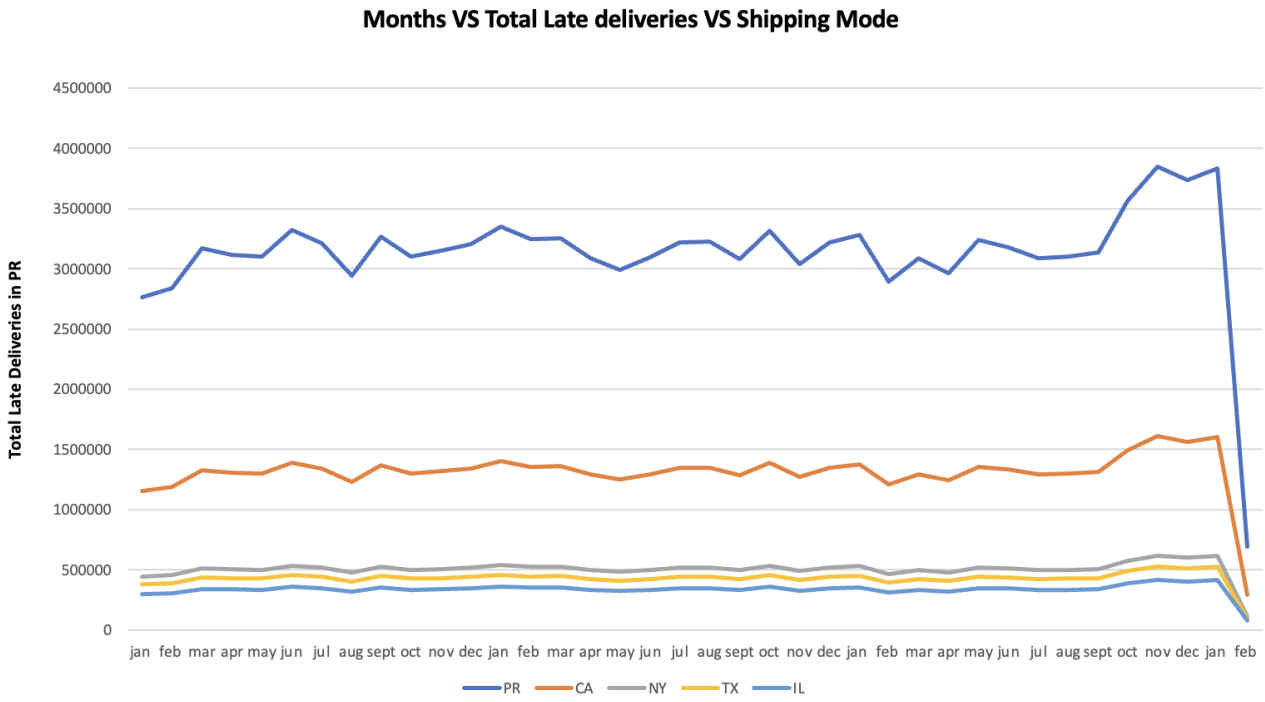
Data Overview: The data reveals that orders shipped from Central America and Western Europe accounted for the highest late delivery count to Puerto Rico.



Task: Examine the correlation between shipping modes and late deliveries to Puerto Rico since we have realized the top 10 regions were overseas to Puerto Rico.

Approach: Our approach in creating this visual was to utilize “Shipping Mode” and “Delivery Status,” with a specific focus on the late delivery count.

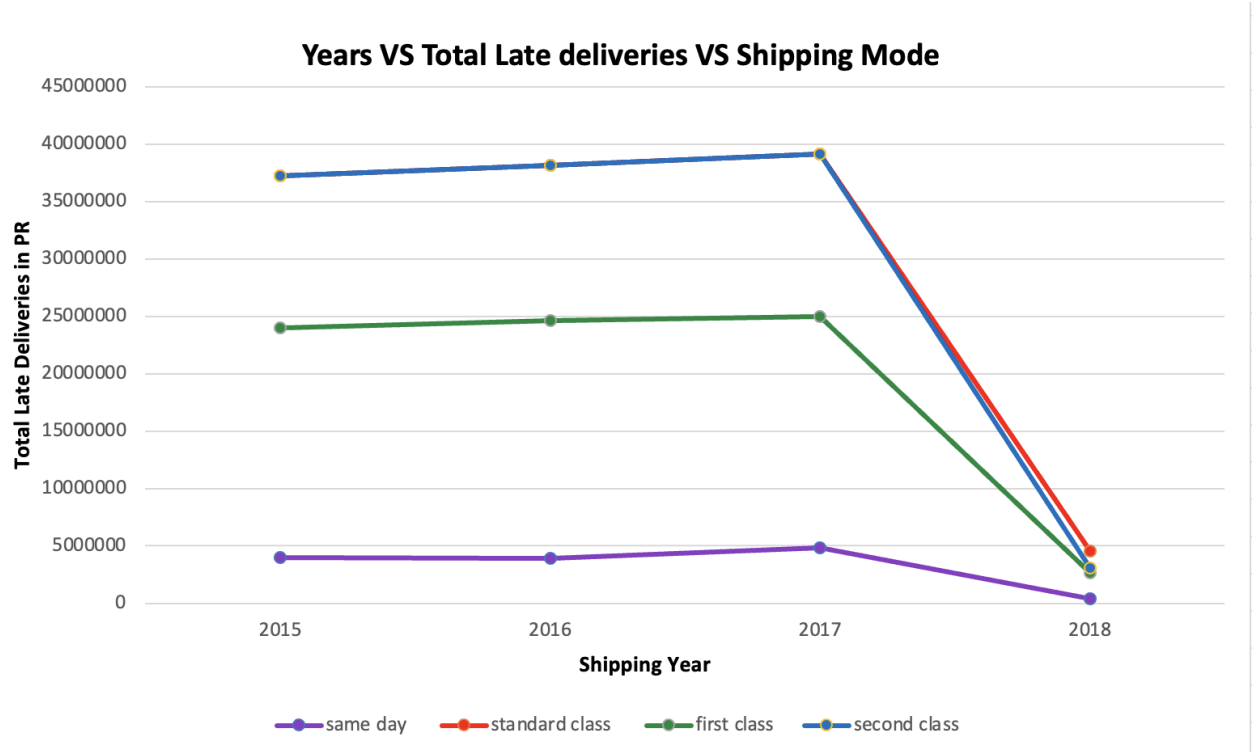
Data Overview: The analysis revealed that standard class had the highest late delivery count accounting for almost 120 Million late deliveries.



Task: Compare late delivery counts between states across all shipping modes overtime by monthly for a span of 3 years; 2015-2018.

Approach: Our approach involved comparing states and examining the variations in late delivery counts. We layered various graphs, running distinct queries, including top 5 “customer states,” total late “delivery counts”, and “shipping year” to make comparisons over time.

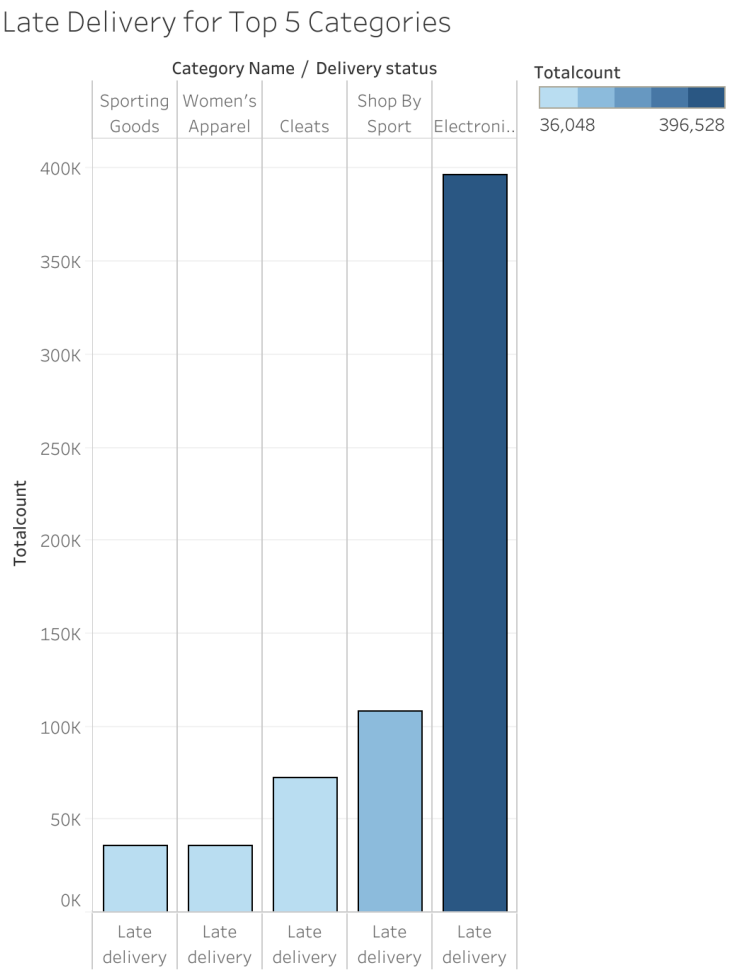
Data Overview: The graph illustrates a significant disparity in late deliveries within Puerto Rico from 2015 to 2018. Despite a decrease over time, late deliveries remain well above the million mark. It is important to note 2018 only has data up to February 6th.



Task: Identify any pattern over time between shipping mode and late deliveries to Puerto Rico over the past years.

Approach: Employing various graphs, we overlaid them to observe the trajectory of each shipping mode concerning late deliveries over the specified period from 2015 to 2018. We utilized data encompassing "Delivery Status," "Total Orders," and "Customer State," with a particular emphasis on late deliveries, shipping modes, and Puerto Rico.

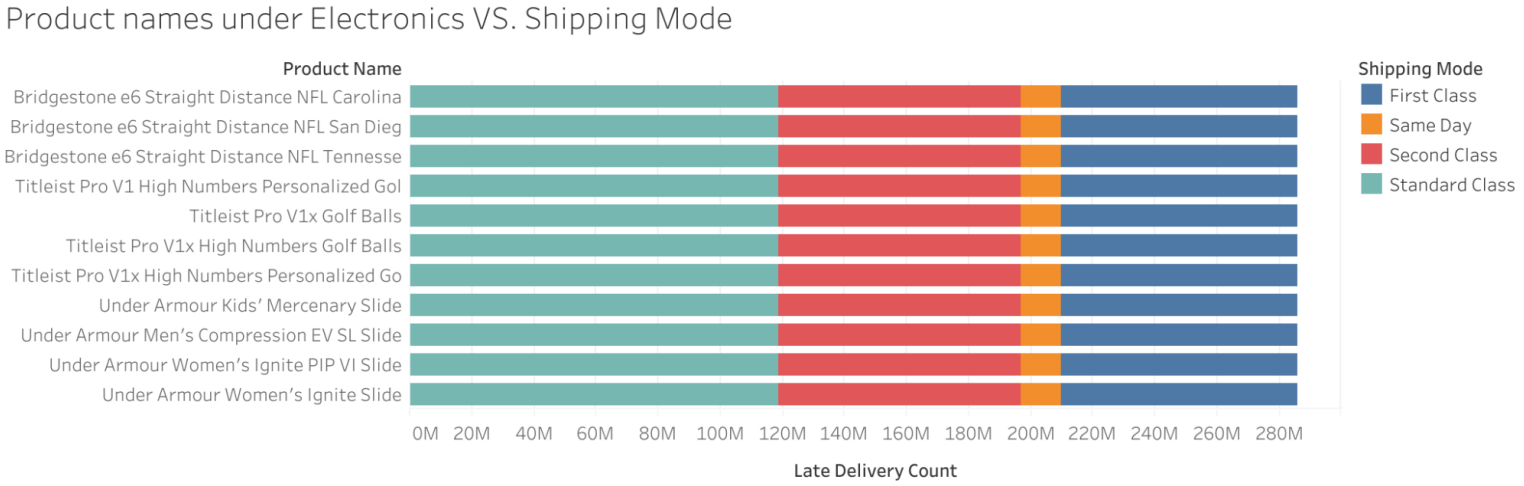
Data Overview: Despite a year-over-year decrease in late deliveries, the number of late deliveries in Puerto Rico remained notably high, surpassing a million in 2018. However, when looking into the data there were only 2 months accounted for in the year of 2018, which is why this number seems to decrease in the graph.



Task: Building on our knowledge so far, our objective was to investigate specific items that were experiencing delayed shipments and explore any potential correlation between late deliveries and these items.

Approach: To identify items experiencing delays, we employed "Delivery Status" and "Category Name," utilizing the Total Count to quantify the occurrences contributing to late deliveries.

Data Overview: Upon conducting an in-depth analysis, it became evident that Electronics constituted the majority of categories associated with late deliveries, as illustrated in the graph above



Task: Identify what products under the Category Electronics were being delivered late in the 4 different shipping modes

Approach: Our approach to creating this visual involved investigating the potential correlation between specific items and late deliveries. We aimed to determine if certain products, such as large TVs, were consistently experiencing delays when shipped to Puerto Rico due to shipping issues. However, the graph creation process using "Electronics," "Shipping Mode," and Total "Late Delivery" Count yielded confusing results, as items like slides and golf balls were appearing under the electronics category.

Data Overview: The data presented unexpected findings, with items not classified as electronics emerging in the results. Further exploration and refinement may be necessary to better understand the correlation between product types and late deliveries.

Conclusion:

Our analysis reveals a notable concern regarding the frequency of late deliveries, particularly when shipping to Puerto Rico. Despite a decline in the number of late deliveries over time, the figures still remain well above the million mark. Additionally 2018 only provides us data up until February 6th, which might also mean that the items could potentially still get delivered late as we don't have the data for the whole month. As we strive to enhance the company's performance this year, we propose the following strategies to mitigate this issue:

1. **Evaluate and Look into Possible Vendors/Suppliers:** Explore partnerships with vendors or suppliers known for their reliability in delivering to Puerto Rico. Prioritize those with access to the region to improve delivery efficiency.
2. **Data Cleaning for Precision:** Conduct a thorough data cleansing process to enhance information accuracy. Specifically, in categories like electronics, ensure that detailed product descriptions are available to avoid confusion, such as distinguishing between slides and tennis balls in the sporting goods category rather than finding them in electronics.
3. **Enhance Tracking Technologies:** Implement advanced tracking technologies to provide customers with real-time updates on their shipments. This improvement in transparency can make customers more likely to come back to shop.
4. **Access to more Customer Registration Data through 2015 - 2018:** No data was provided regarding customers such as returning/repeating customers or customer experience. We can perform additional analysis with this data to learn whether customer count has increased or decreased through the years due to late delivery.

- **Access to size and weight of the products/items under specific categories**, can also help in additional analysis, as the weight of certain items can lead to added shipping time, cost, space, security and processing time.

As previously mentioned, by addressing these pain points, we anticipate not only enhancing the customer experience but also positively impacting the company's performance metrics. Improving the reliability of deliveries and ensuring adequate stock levels will contribute to customers returning to repeat business, and ultimately, the overall success of the company.

Works Cited

All data and information was pulled from John Droeschers's database on "JohnDroescher.com"