Analyzing Late Deliveries in Shipping

Paulette Rodriguez

March 2025

Table of Contents

[Objective 1](#_Toc194003248)

[The Data Analysis 2](#_Toc194003250)

[Conclusion: 9](#_Toc194003251)

## Objective

Imagine you’re a data analyst intern at a retail company that sells products across different regions. The company wants to understand why some of their deliveries are late and how they can improve their shipping process. For this project, our goal is to analyze the data about customer orders, shipping methods, and product information to find patterns that could explain the risk of late deliveries and help the company make better decisions.

For the project will be working with the three different data sets:

1. **Order Table**: This data set contains information about the orders, including the date, shipping mode and whether the other is a risk for late delivery.
2. **Customer Table**: This data set contains customer information such as their name, city, and state.
3. **Product Table**: This data set contain product details such as the product names, category, and price.

Our goal for this project is analyze the corresponding data sets in Microsoft Excel to provide useful insights to the company and answer the following questions:

* Does the shipping mode affect the risk of late deliveries?
* Are certain regions more likely to have late deliveries?
* Are certain product categories more likely to be associate with late deliveries?
* Who are the top customers with the latest deliveries?
* What percentage of total orders are at risk for late delivery?

Our objective is to perform a basic data analysis on the data set to answer the questions listed above. With the use of Excel, we will perform a step-by-step analysis to understand and prepare the data, clean the data and answer the key questions using visualization tools such as charts and pivot tables.

## The Data Analysis

#### Importing & Cleaning Data

First, we will begin by importing and cleaning the data set. This will help make sure the data set is running properly and any mistakes are fixed before we begin the data analysis. After importing the data, we take care of cleaning the data and making sure that the columns we will need are in the correct format. For example, after importing we want to first check for any missing data. Looking for any missing values (blank customer names or product details) will help making working with the data set easier and much more efficient. Then, we want to check formats to ensure that things like dates are correctly formatted. Lastly, combining the data using VLOOKUP will help bring together related data. For example, we may want to use a VLOOKUP to link Customer ID from the Order Tabel to the Customer ID in the Customer Table and link Product ID from the Order Table to the Product ID in the Product table. This will allow us to see customer details (like name and city) alongside their orders.

After preparing the data, we can begin to analyze it to answer the following questions. With tools such as Pivot Tables and Charts, we will then summarize our findings and visualize the data.

#### Question 1: Does the shipping mode affect the risk of late deliveries?

Our task for this first question is to analyze how the different shipping modes (e.g., Second Class, First Class) relate to the late delivery risk of the orders being purchased. We begin by creating a pivot table with Shipping Mode in rows and Late\_Delivery\_Risk in the values.

|  |  |
| --- | --- |
| **Shipping Mode** | **Late Delivery Risk** |
| First Class | 73 |
| Same Day | 13 |
| Second Class | 297 |
| Standard Class | 246 |
| **Grand Total** | **629** |

Provided the table, we can now check to see which of the different shipping modes is associate with more late deliveries.

Given the pivot table above, and the provided output, we can conclude that some shipping methos are riskier than others. For example, Second Class shipping mode comes in as the mode with the most late deliveries, with Standard Class shipping mode coming in second. By using bar chart and a pie chart we can visualize these results.

A pie chart helps us in showing the proportion of on-time and late deliveries for each shipping mode. Given our analysis, we can conclude that the overall 47% of packages being delivered by Second Class shipping mode arrive late to their destination.

The given percentage of late deliveries may be due to many different factors including, but not limited to, location, regions, size of the package, seasonal occurrences that may disrupt the overall shipping process of certain items. In a more general sense, the shipping modes may also vary in pricing. The company may want to think about investing in pricing options that are more suitable to the types of customers buying the products. In these cases, the company may want to invest in a better shipping mode so their products not only get delivered to their customers on time, but so their customer satisfactory rate could be improved.

#### Question 2: Are certain regions more likely to have late deliveries?

Our task for the second question is to investigate if late deliveries are more common in specific regions (e.g., West Africa, Central Africa). We begin by creating a pivot table with Order Regions in the rows and Late\_Delivery\_Risk in the values. This will allow us to compare the late deliveries in different regions.

|  |  |
| --- | --- |
| **Regions** | **Late Delivery Risk** |
| Canada | 2 |
| Caribbean | 16 |
| Central Africa | 22 |
| Central America | 56 |
| Central Asia | 1 |
| East Africa | 25 |
| East of USA | 27 |
| Eastern Asia | 33 |
| Eastern Europe | 10 |
| North Africa | 35 |
| Northern Europe | 44 |
| Oceania | 25 |
| South America | 38 |
| South Asia | 36 |
| South of USA | 13 |
| Southeast Asia | 29 |
| Southern Africa | 19 |
| Southern Europe | 26 |
| US Center | 18 |
| West Africa | 44 |
| West Asia | 5 |
| West of USA | 22 |
| Western Europe | 83 |
| **Grand Total** | **629** |

Provided the pivot table output, we can now compare the late deliveries in different regions and visualize these results using a bar chart.

From this chart, we can conclude a couple of things. Overall, Western Europe has the highest number of late deliveries, with Central America coming in as second highest in late deliveries. West Africa, Northern Europe and South America are amongst the top five regions with the highest risk of late delivery. This bar chart is very useful and helps us visually capture all the regions included within our data set.

Working with smaller sets might be beneficial as it allows us to filter this data for different regions around the world. If we wanted to view on the regions in the United States, we would be able to do that as well. As an example, we will now focus on African regions and filter the pivot table to show only regions in Africa such as, Central Africa, East Africa, North Africa, Southern Africa, and West Africa.

|  |  |
| --- | --- |
| **Regions** | **Late Delivery Risk** |
| Central Africa | 22 |
| East Africa | 25 |
| North Africa | 35 |
| Southern Africa | 19 |
| West Africa | 44 |
| **Grand Total** | **145** |

If we filter by region, we find that certain regions, West Africa and North Africa, have more late deliveries than the others.

In general, this could indicate regional issues with logistics and shipping reliability for the company. As a result, the company can think of ways to improve their delivery services, cut transportation costs, strengthen communication at all levels, establish standard suppliers and partners, invest in the right technology and reduce warehouse management errors. These are just some of the ways the company can begin to transform and fix their late delivery issues.

#### Question 3: Are certain product categories more likely to be associate with late deliveries?

Our task for the third question is to see if some product categories are linked to higher late delivery risks as compared to other products being sold. We first begin by creating a pivot table with Product Category Name in the rows and Late\_Delivery\_Risk in the values. This will allow us to compare the late deliveries in for all the products available to us in the products table.

|  |  |
| --- | --- |
| **Product Categories** | **Late Delivery Risk** |
| Accessories | 9 |
| As Seen on TV! | 1 |
| Baseball & Softball | 3 |
| Basketball | 1 |
| Boxing & MMA | 2 |
| Cameras | 10 |
| Camping & Hiking | 14 |
| Cardio Equipment | 60 |
| Children's Clothing | 2 |
| Cleats | 114 |
| Computers | 4 |
| Consumer Electronics | 11 |
| Crafts | 3 |
| Electronics | 19 |
| Fitness Accessories | 3 |
| Girls' Apparel | 3 |
| Golf Balls | 3 |
| Hockey | 42 |
| Hunting & Shooting | 0 |
| Kids' Golf Clubs | 1 |
| Lacrosse | 0 |
| Men's Clothing | 0 |
| Men's Footwear | 95 |
| Shop By Sport | 41 |
| Soccer | 1 |
| Sporting Goods | 85 |
| Strength Training | 1 |
| Tennis & Racquet | 2 |
| Trade-In | 11 |
| Women's Apparel | 85 |
| Women's Clothing | 3 |
| **Grand Total** | **629** |

Provided the pivot table output, we can now compare the risk of late deliveries by the product categories using a bar chart.

For example, we want to analyze if products such as Electronics or Cardio Equipment tend to have more late deliveries. We might discover that many of the cardio and sporting equipment tend to arrive later. We can conclude that this may be due to their size, complexity, shipping requirements, or the location these packages are being sent to.

#### Question 4: Who are the top customers with the most late deliveries?

For question four, our task is to identify the customer who have experienced the most late deliveries. First, we begin by creating a pivot table with Customer details in rows and Late\_Delivery\_Risk in values. By doing this analysis, we will be able to count how many late deliveries each customer has had. We will use a bar chart to show the results and compare the customers with the most late deliveries.

Our result concludes that one customer out of the rest has had the highest number of late deliveries. The top customer with the highest number of late deliveries is Mary Smith, with a total of 89 late deliveries. This result indicates that there are major shipping issues that are affecting their orders. This result helps the company prioritize customer service for this customer or begin to investigate possible shipping issues. One of the main concerns may be focused on if this customer is receiving the best shipping mode the company can offer. A goal the company may want to strive for is making sure this customer is being taken care of with the best shipping mode they can offer to try and expedite the order to them in the future.

#### Question 5: What percentage of total orders are at risk for late delivery?

The task for this question is to find out what proportion of the orders are at risk of being delivered late. We begin by first counting the total number of orders in the Order Table. Then, we count how many orders have a Late\_Delivery\_risk equal to 1 (indicating a late risk). Then, we want to finish off by calculating the percentage of late orders as following:

This percentage will give us an idea of how big the problem is. For example, if 20% of all orders area at risk of being late, then the company may need to take action to improve shipping processes.

For this specific data set, by calculating the percentage of late order using the formula above, we get the following result:

As a result, we conclude that about 49.6% of orders are at risk of late delivery. This is a pretty high percentage of late deliveries and something the company should be very worried about. The company should take immediate action to improve their shipping processes provided the previous analysis we have done on shipping modes. Overall, if the company is shipping to far regions, they may want to focus on switching carriers that will benefit those who purchase outside of their territory as compared to those who purchase domestically.

## Conclusion:

For this project, our goal was to analyze the data about customer orders, shipping methods, and product information to find patterns that could explain the risk of late deliveries and help the company make better decisions. Given the data analysis performed we were able to explore patterns in late deliveries and answer the following questions:

* Does the shipping mode affect the risk of late deliveries?

Per the analysis performed, we were able to conclude that the Second Class shipping mode causes the highest volume of late deliveries. Given our analysis, we were able to conclude that overall, 47% of packages being delivered by Second Class shipping mode arrive late to their destination. Some recommendations for the company to follow in such cases is to figure out what is causing the lateness and have open communication with their carrier. Many times, shipping status is due to errors being made in the warehouse. One thing the company can strive to do is make sure their carriers and warehouse employees have a small margin of errors when preparing orders to be shipped.

* Are certain regions more likely to have late deliveries?

Overall, we were able to conclude that Western Europe has the highest number of late deliveries, with Central America coming in as second highest in late deliveries. West Africa, Northern Europe and South America are amongst the top five regions with the highest risk of late delivery. As an example, we focused on a smaller set related to only African regions and filtered the pivot table to show only regions in Africa such as, Central Africa, East Africa, North Africa, Southern Africa, and West Africa. After filtering by region, we were able to conclude that certain regions, West Africa and North Africa, have more late deliveries than the others.

In general, this could indicate regional issues with logistics and shipping reliability for the company. As a result, it is recommended that the company think of ways to improve their delivery services, cut transportation costs, strengthen communication at all levels, establish standard suppliers and partners, invest in the right technology and reduce warehouse management errors.

* Are certain product categories more likely to be associate with late deliveries?

Given the performed analysis, we were able to conclude that the cardio and sport equipment are more likely to be associated with late deliveries. This may be due to their size and complexity. The company is recommended to work through these issues by re-evaluating how products in certain categories are shipped. Perhaps, the company and carrier need to more time to handle these types of products before shipping.

* Who are the top customers with the most late deliveries?

We were able to conclude that one customer, Mary Smith, is the customer with the most late deliveries out of the entire data set. The company is recommended to offer more proactive customer service to customer who have frequent late deliveries by offering better shipping modes, and different carriers dependent on location and time.

* What percentage of total orders are at risk for late delivery?

Based on the available data and the calculations made, we were able to conclude that a high volume of orders, 49.6% of orders, are at risk for late delivery. The company may want to try switching to faster shipping methods in certain regions to reduce late deliveries.

By completing this analysis, we were able to provide the company with actionable insights to improve shipping reliability and customer satisfaction. These insights also allowed us to work on a data set to showcase our data analysis skills in a portfolio or future employers. This project allowed us the opportunity to showcase our ability to organize and clean data, use pivot tables and charts to find meaningful insights, and present our findings concisely.