Exploratory Analytics and Data Wrangling For E-Commerce Analytics

Executive Report

Caio di Felice Cunha

Dear Executive Board,

I hope this email finds you well.

Below you will find valuable insights about our company.
For more details, please visit our interactive dashboard

Conclusions:

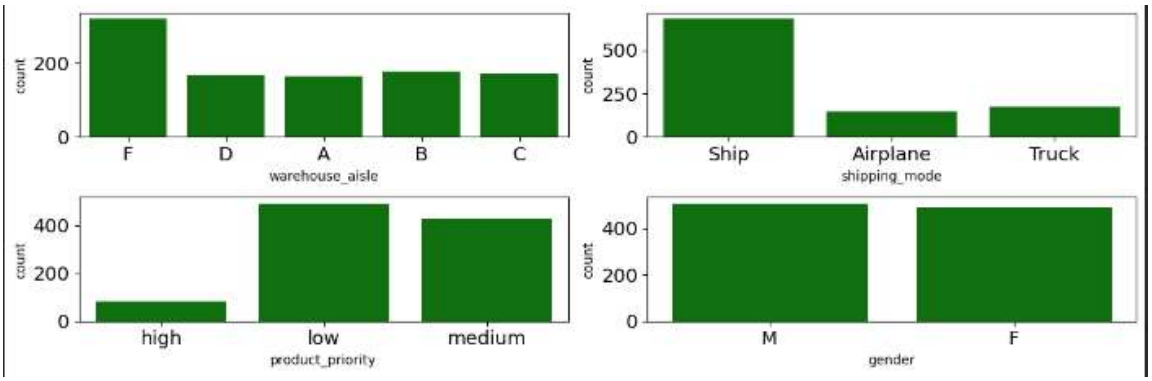
It was concluded that there are several products with delay issues.
In several areas (shipping mode, product priority, among others) products are experiencing a delay of 50% or more.
Therefore, it is necessary to better understand the transport market in order to hire better companies that meet the deadline for delivering products.

Analysis Summary

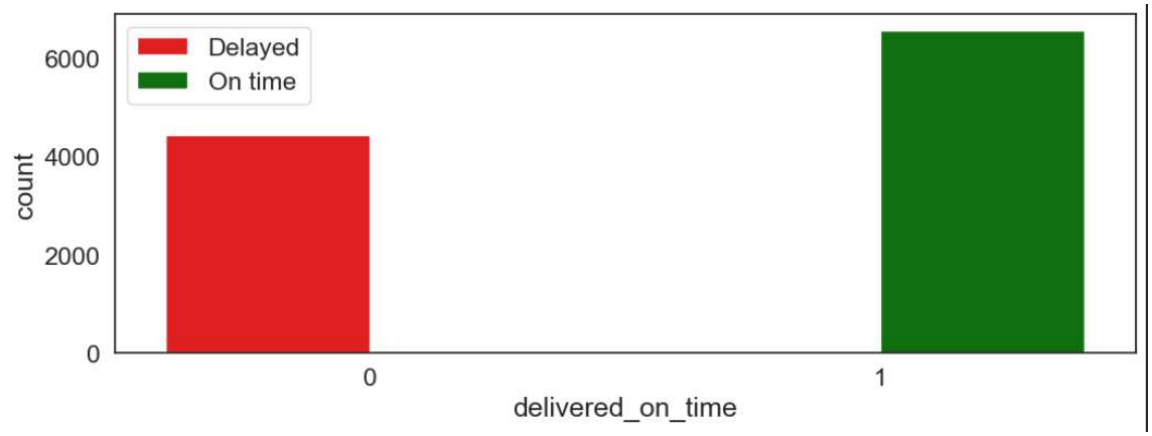
General observations:

1. Most of the products shipped came from aisle **F.**
   1.1 Total Records By Variable Category warehouse_aisle:
   * F 3,666
   * D 1,834
   * C 1,834
   * A 1,833
   * B 1,833
2. The most used shipping method is **Ship**.
   2.1 Total Records By Variable Category shipping_mode:
   * Ship 7,463
   * Airplane 1,777
   * Truck 1,760
3. The priority of purchased products are **low** category.
   3.1 Total Records By Variable Category product_priority:
   * low 5,298
   * medium 4,754
   * high 948
4. Dealing with gender, **Female** is the one that appears the most.
   4.1 Total Records By Variable Category gender:

    * F 5,546
    * M 5,454



When we look at products that were delivered late or on time:



From the graphs it can be seen that:

- Shipping mode via Ship is most common.
- There are few products with high delivery priority.
- The proportion of Male and Female customers is practically the same.
- We have more on-time deliveries, however the number of late deliveries seems to be at a high level.

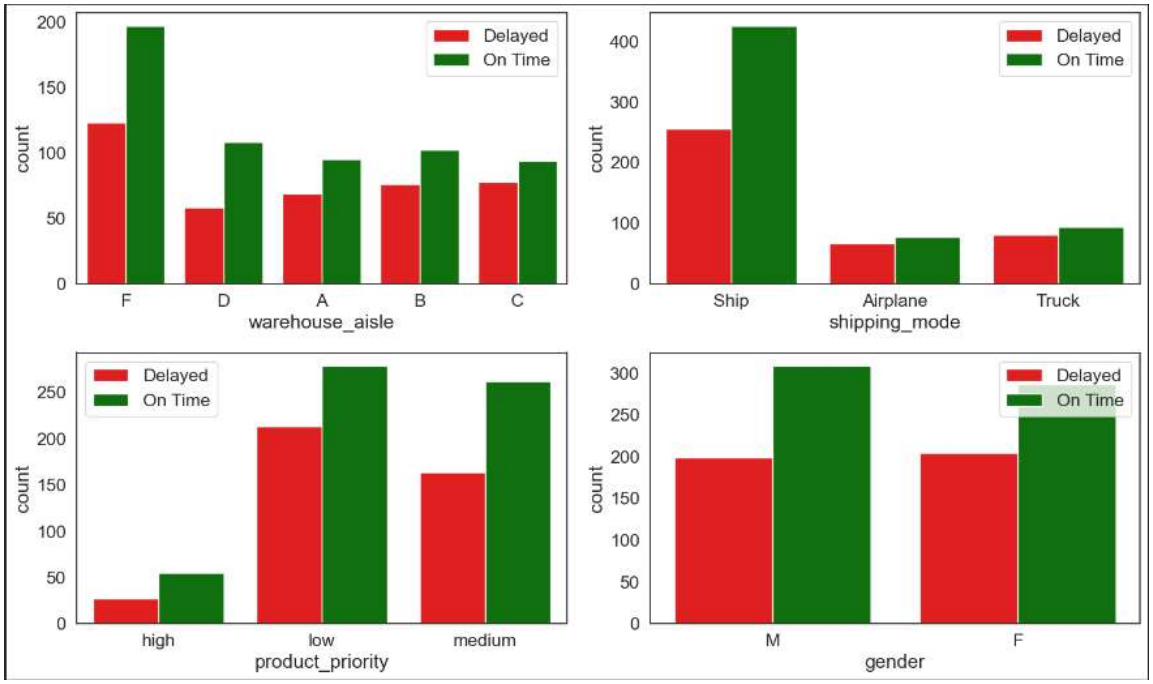
From the above *correlation analysis* it can be seen that:

- There is a positive correlation between `number_customer_calls` and `product_cost`.
- There is a positive correlation between `number_customer_calls` and `previous_purchases`.
- There is a negative correlation between `discount` and `weight_grams`.

When looking at the graph below, we see:

- Shipping by ship tends to be late in delivery.

- In priority of the product the low and medium categories tend to have delays in delivery.
- In warehouse aisle, products from corridor F tend to be late in delivery.
- Products purchased by women have a slightly lower volume of on-time deliveries.

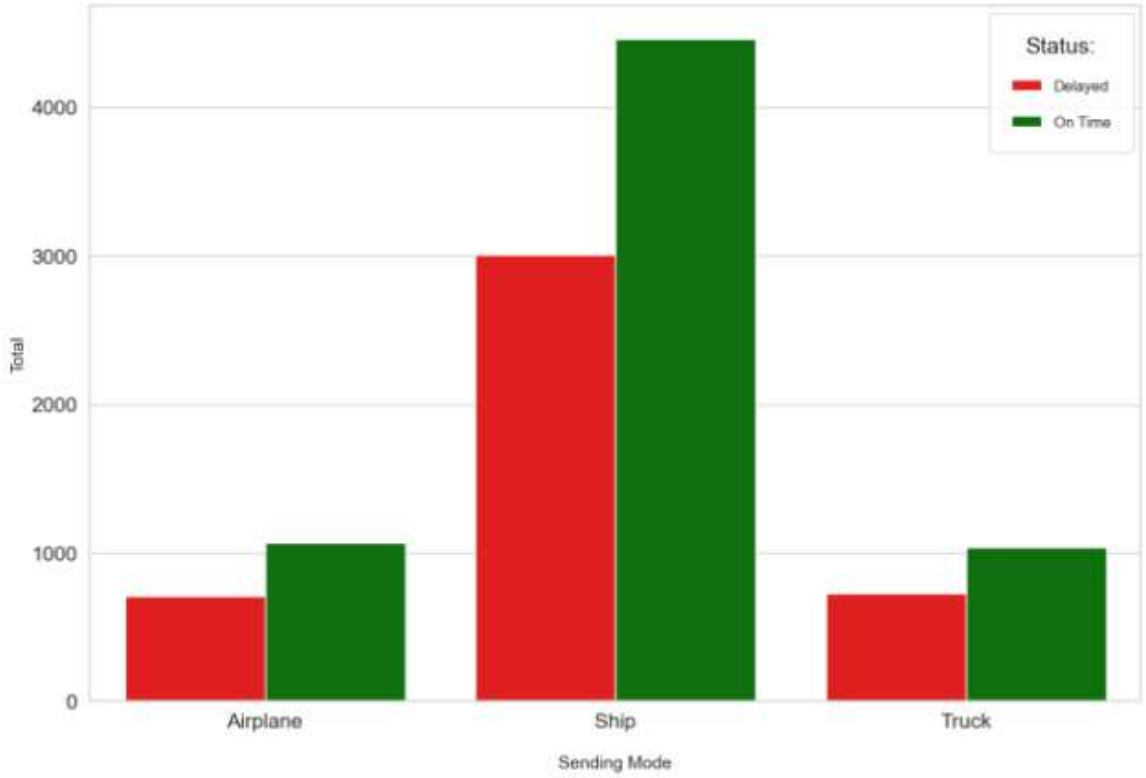


Shipping mode:

	Deliver Status	Shipping Mode	Total per Deliver Status per Shipping Mode	Total per Deliver Status
0	Delayed	Airplane	708	4437
1	Delayed	Ship	3004	4437
2	Delayed	Truck	725	4437
3	On time	Airplane	1069	6563
4	On time	Ship	4459	6563
5	On time	Truck	1035	6563

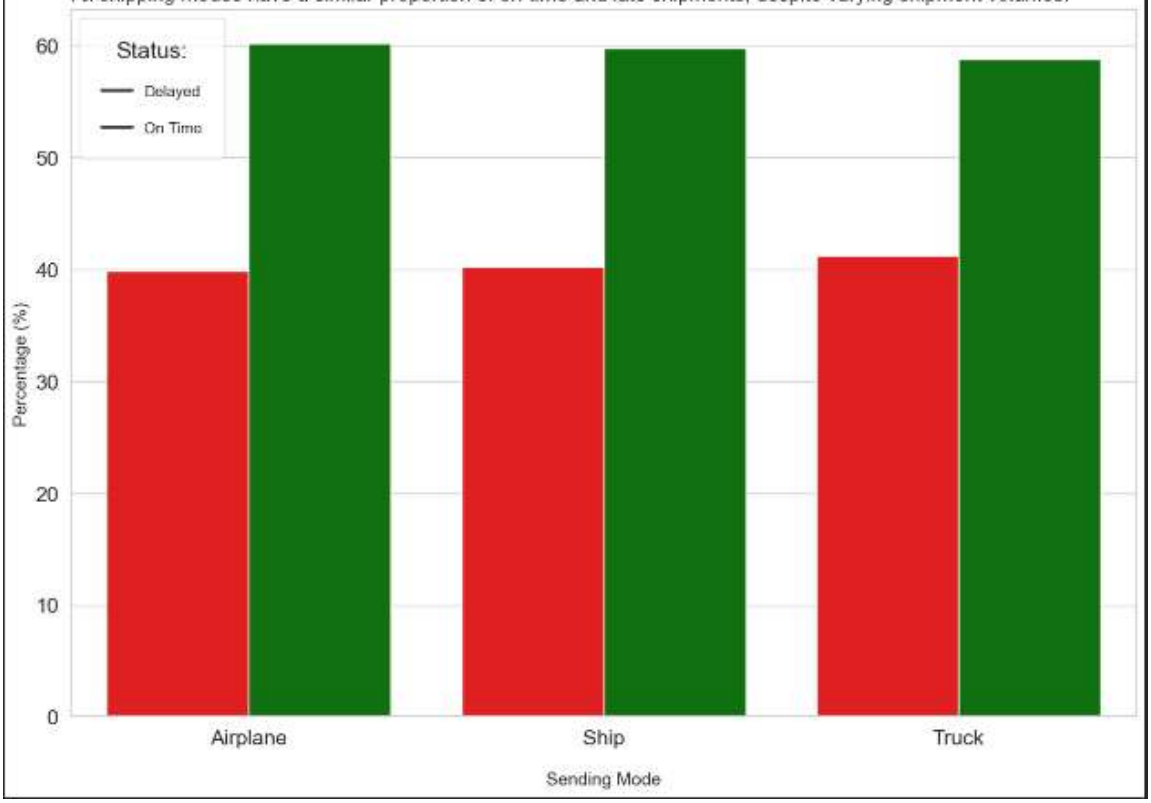
Deliveries Based on Shipping Mode (Absolute)

There are delays with all shipping modes, but using Ship seems to be the mode that causes the most delays.



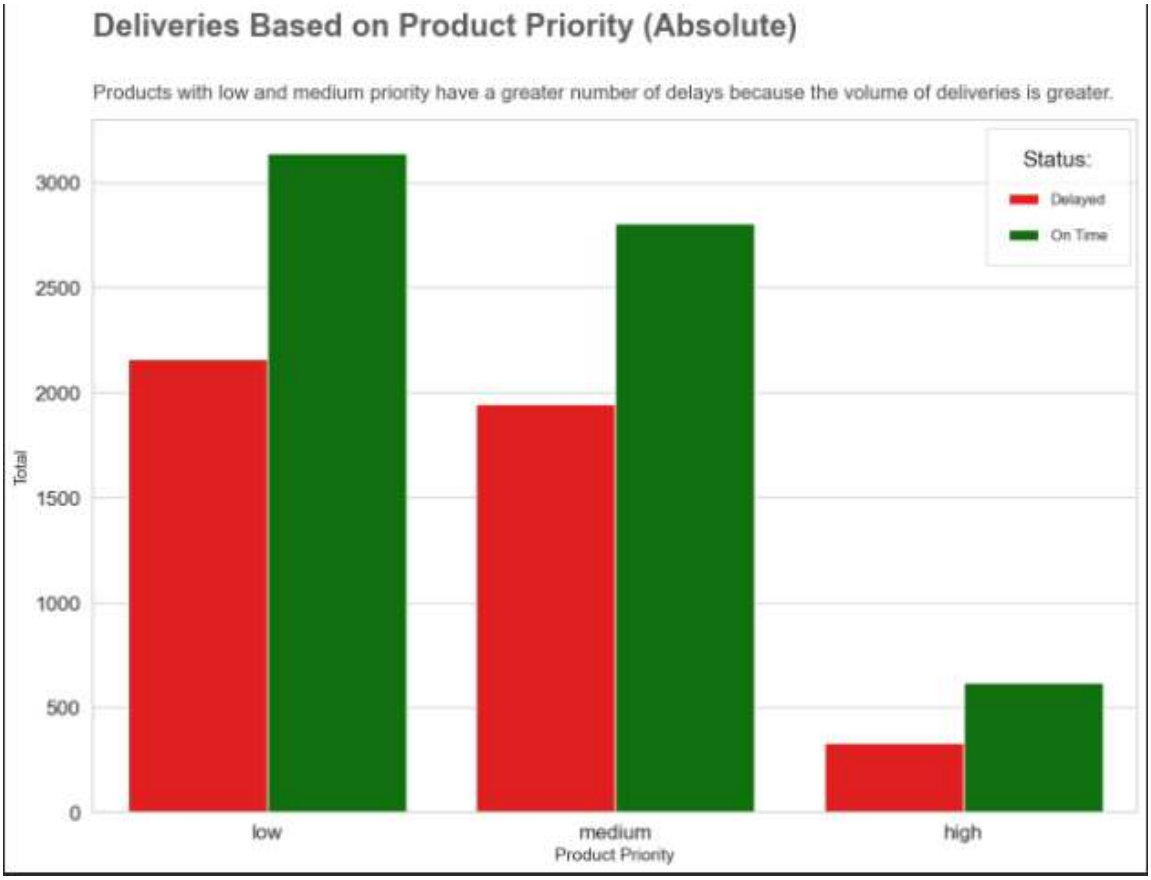
Deliveries Based on Shipping Mode (Percentage)

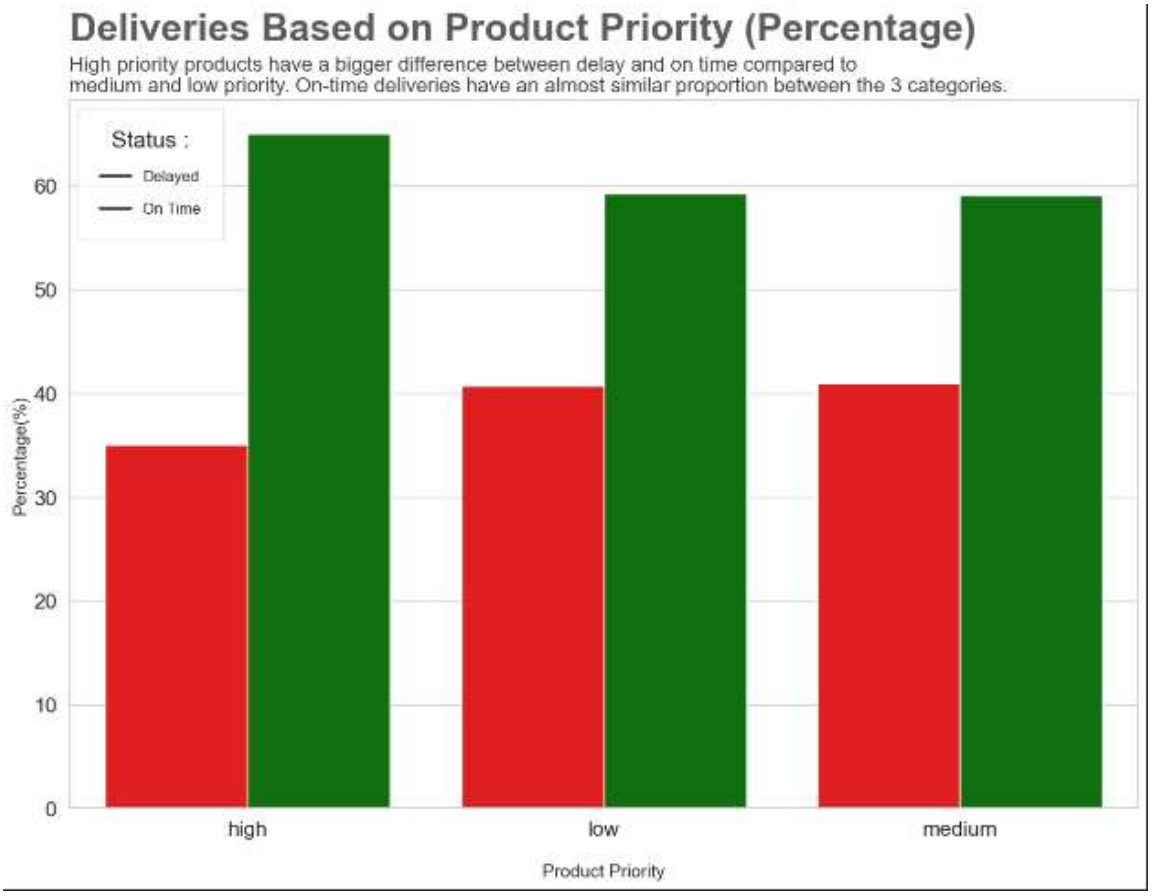
All shipping modes have a similar proportion of on-time and late shipments, despite varying shipment volumes.



Product Priority

	Deliver Status	Priority of the Product	Total per Deliver Status per Product Priority	Total per Deliver Status
0	Delayed	high	332	4437
1	Delayed	low	2158	4437
2	Delayed	medium	1947	4437
3	On Time	high	616	6563
4	On Time	low	3140	6563
5	On Time	medium	2807	6563





The Influence of the Discount over the Deliver Status of products

Thank you for taking the time to read my message, and I look forward to hearing back from you soon.

Best regards,

Caio di Felice Cunha

End