ITE-5201 – Fall 2022

Humber College IGS

Final Project Report

Introduction to Data Analytics

Project Title:

Visualization / Analysis of

E-Commerce Shipping Data

Prepared By:

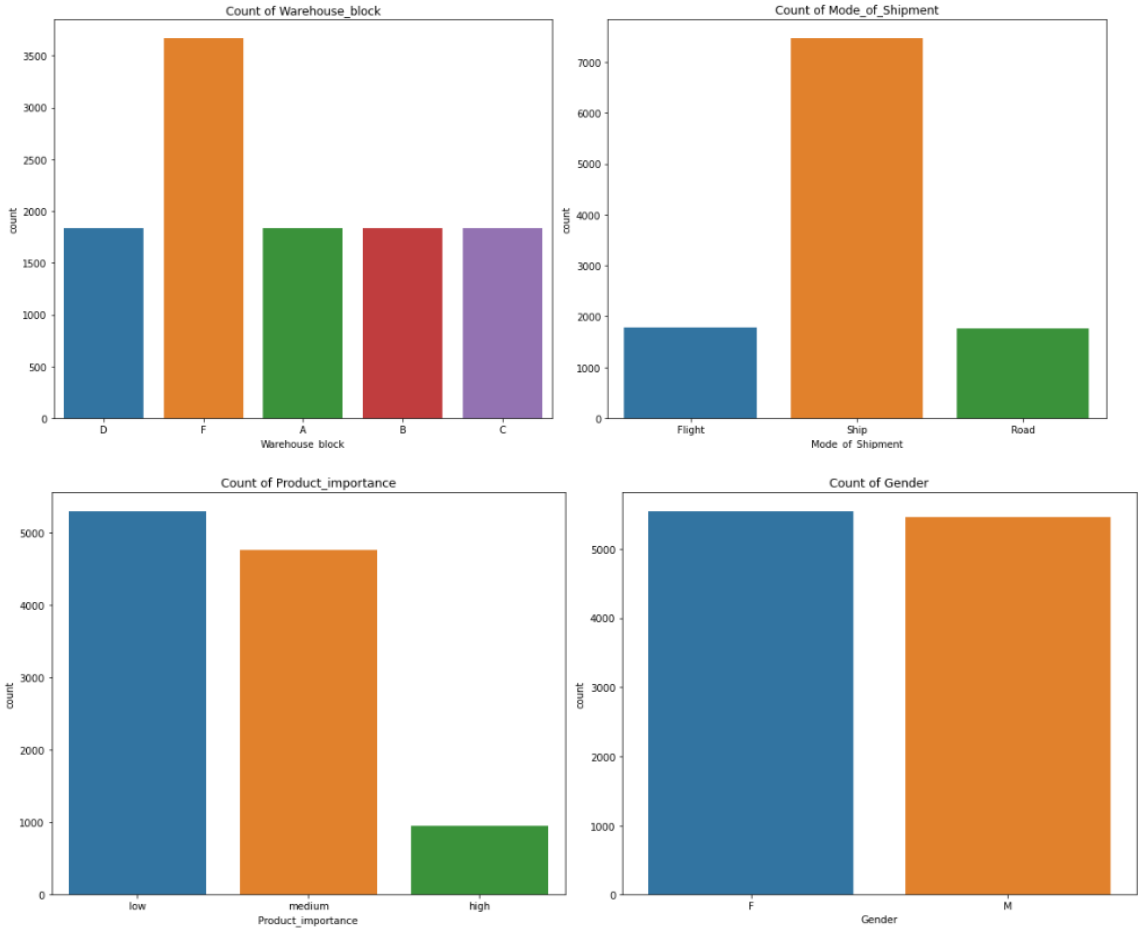
Meet Trivedi (N01520331)

Raghavender Chakilam (N01533573)

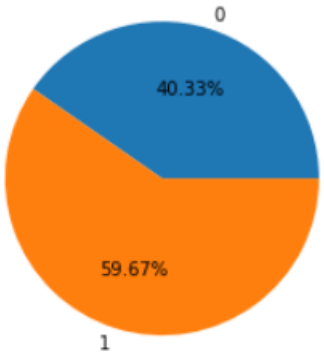
1. Problem statement:
   ⇒ Visualization / Analysis of eCommerce shipping data.

2. Dataset description:
   ⇒ The E-commerce shipping dataset contains daily shipping logs of good which was delivered from different warehouse to customers.
   ⇒ Dependent variables such as warehouse_block, mode_of_shipment, customer_care_calls, customer_rating, cost_of_the_product, prior_purchases, product_importance, discount_offered, gender and weight_of_goods, which we have used to predict if goods received by customer were on time or not.

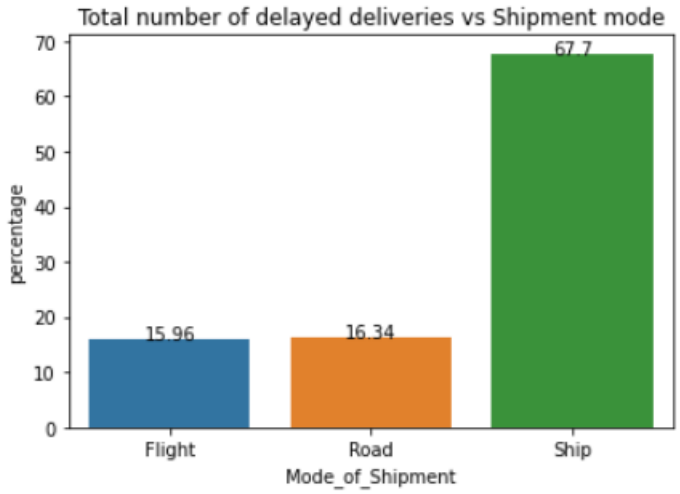
3. Dataset analysis and observation:
   ⇒ For dataset analysis, we have used heatmap and subplot for finding correlation coefficient rank and for observation using above columns.



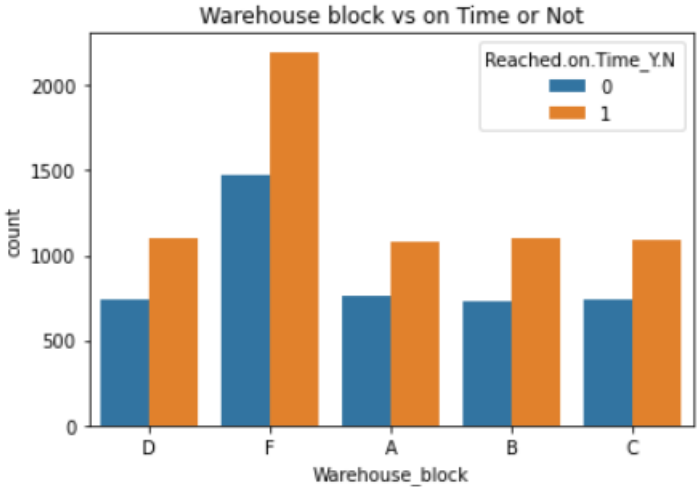
Observation: Above graphs show count for various independent variables.



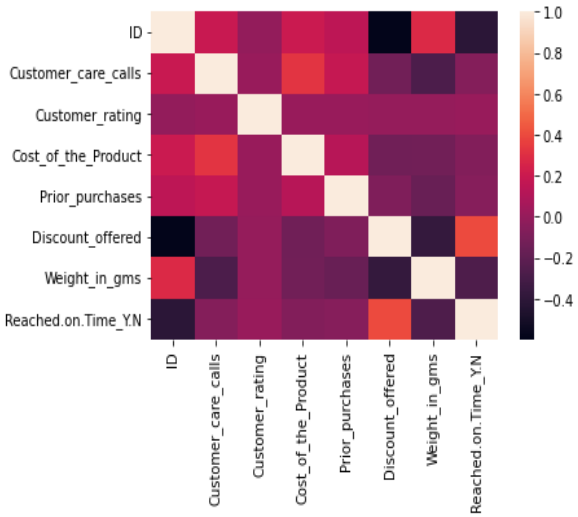
Observation: 60% of the goods were not delivered on time



Observation: About 68% of the delayed deliveries were because of ships mode of delivery



Observation: Above plot shows number of deliveries for each warehouse



4. Proposed Analytical Model:
    - ⇒ Since the target variable has only two outcomes, we have used logistic regression.
    - ⇒ Depending on various dependent variables we predicted the outcome whether the goods are delivered on time or not.
5. Result and Discussion:
    - ⇒ From the below classification report, the model has a precision of 54% and recall of 57%, and accuracy of 64%.

	precision	recall	f1-score	support
0	0.54	0.57	0.56	1313
1	0.71	0.68	0.69	1987
accuracy			0.64	3300
macro avg	0.63	0.63	0.63	3300
weighted avg	0.64	0.64	0.64	3300