Neeman’s Analysis Report

import pandas as pd  
import numpy as np  
import warnings  
warnings.simplefilter('ignore')

data = pd.read\_csv('D:/Datasets/Assignment\_Revenue&Exchanges (3).csv', encoding\_errors= 'replace', low\_memory= False)

data.head(20)

data['Order Amount'].value\_counts().head(50)

data.dtypes

data['Order Amount'] = data['Order Amount'].replace(' - ', '0')

data['Order Amount'].value\_counts().head(10)

def convert\_string(str):  
 num = str.replace(',' , '')  
 return int(num)

Created a function to convert string data type to integer data type.

convert\_string(data['Order Amount'][0])

Order\_Amount = [convert\_string(i) for i in data['Order Amount']]

Using list comprehensions created a list of order price amounts

data['Order Amount'] = pd.Series(Order\_Amount)

data

data.dtypes

data['Ship PinCode'][0]

data['Ship PinCode'].value\_counts().head(80)

data['Ship PinCode'] = pd.to\_numeric(data['Ship PinCode'], errors= 'coerce')

Converted ship pincode object datatype to numeric datatype.

data.dtypes

data['Unit Price'] = data['Unit Price'].replace(' - ', '0')

unit\_price = [convert\_string(i) for i in data['Unit Price']]

Using list comprehensions created a list of unit price amounts

data['Unit Price'] = pd.Series(unit\_price)

data.dtypes

data['Discount'] = data['Discount'].replace(' - ', '0')

Discount = [float(i.replace(',','')) for i in data['Discount']]

data['Discount'] = pd.Series(Discount)

data.dtypes

data.head(15)

pd.to\_numeric(data['Ship PinCode'],downcast= 'signed')

def convet\_integer(float):  
 num = int(float)  
 return num

data['Ship PinCode'] = data['Ship PinCode'].replace(np.nan, 0)

data['Ship PinCode'] = data['Ship PinCode'].apply(convet\_integer)

Using apply function to convert the data type of feature column.

data.dtypes

data['Size'] = data['Size'].replace(np.nan, 0)

data['Size'] = data['Size'].apply(convet\_integer)

data['Order Qty'] = data['Order Qty'].replace(np.nan, 0)  
data['Shipped Qty'] = data['Shipped Qty'].replace(np.nan, 0)  
data['Order Qty'] = data['Order Qty'].apply(convet\_integer)  
data['Shipped Qty'] = data['Shipped Qty'].apply(convet\_integer)

data.head(15)

data.to\_excel('Neeman\_data.xlsx')

Created a excel with the clean and transformed data.

data['Ship State'].value\_counts().head(10)

Checking the top 10 shipping locations of the products.

data['Discount'].describe()

discount = data['Discount'].values

max(discount)

np.where(discount == max(discount))

data['Discount'][160154]

data = data.drop(index= [160145, 160146, 160147, 160148, 160149, 160150, 160151, 160152,  
 160153, 160154, 160155, 160156, 160157, 160158, 160159, 160160,  
 160161, 160162, 160163, 160164, 160165, 160166, 160167, 160168,  
 224575, 224576, 224577, 224578, 224579, 224580, 224581, 224582,  
 224583, 224584, 224585, 224586, 224587, 224588, 224589, 224590,  
 224591, 224592, 224593, 224594, 224595, 224596, 224597, 224598], axis= 0)

As the discount for the above records is 145154 for one product and that cannot be possible so dropped those records.

data

data['City'].value\_counts().head(20)

Top 10 shipping cities

data['City'] = data['City'].replace('MUMBAI', 'Mumbai')

As the dataset having two same cities with a different spelling so replaced with the single one.

data['City'] = data['City'].replace('PUNE', 'Pune')

data['City'].value\_counts().head(20)

filter\_maharastra = data[data['Ship State'] == 'Maharashtra']

Created a data frame where the shipping location is only Maharashtra.

filter\_maharastra.describe()

Checked the statistical analysis of data frame.

filter\_maharastra['City'].value\_counts().head(30)

Top 30 shipping cities in Maharashtra.

filter\_maharastra['City'] = filter\_maharastra['City'].replace('THANE', 'Thane')  
filter\_maharastra['City'] = filter\_maharastra['City'].replace('NAGPUR', 'Nagpur')  
filter\_maharastra['City'] = filter\_maharastra['City'].replace('NAGPUR', 'Nagpur')  
filter\_maharastra['City'] = filter\_maharastra['City'].replace(['NASHIK','Nasik'], 'Nashik')  
filter\_maharastra['City'] = filter\_maharastra['City'].replace('mumbai', 'Mumbai')  
filter\_maharastra['City'] = filter\_maharastra['City'].replace('pune', 'Pune')

filter\_maharastra['City'].value\_counts()

data['Size'].value\_counts()

Variants of Product Distribution.

data['Status'].value\_counts()

Status of the orders distribution

filtered\_delivered = data[data['Status'] == 'Delivered']

Created a data frame where the status of the orders is delivered.

filtered\_delivered

filtered\_delivered['Size'].value\_counts().

filtered\_delivered['Ship State'].value\_counts()

data.head(10)

data['OrderDate'] = pd.to\_datetime(data['OrderDate'], format='%d-%m-%Y')

Changed the order date into above specified format.

data.dtypes

Checked the data types of the features.

data['OrderMonth'] = data['OrderDate'].dt.month

Separated the order date and created an order month.

data['OrderYear'] = data['OrderDate'].dt.year

Separated the order date and created an order year.

data.head(10)

data['OrderMonth'].value\_counts()

filtered\_delivered['OrderDate'] = pd.to\_datetime(filtered\_delivered['OrderDate'], format='%d-%m-%Y')

filtered\_delivered['OrderMonth'] = filtered\_delivered['OrderDate'].dt.month  
filtered\_delivered['OrderYear'] = filtered\_delivered['OrderDate'].dt.year

filtered\_delivered['OrderMonth'].value\_counts()

Monthly wise delivered products.

filtered\_delivered

data.groupby('Status')['Discount'].sum()

data.groupby('Order Amount')['Discount'].count().sort\_values(ascending= False)

Grouped the order amount data with the counts of discount

data['Order Amount'].value\_counts()

data[data['Order Amount'] == 2]

filtered\_delivered.groupby('Order Amount')['Discount'].sum().sort\_values(ascending= False).head(30)

Grouped the order amount data with the sum of discount amounts.

filtered\_delivered[filtered\_delivered['Order Amount'] == 2]

filtered\_delivered

filtered\_delivered = filtered\_delivered.drop(filtered\_delivered[filtered\_delivered['Order Amount'] == 2].index)

I see that there is a large sum of discount amounts for the order amount 2 so I have gone ahead and dropped those records.

filtered\_delivered.groupby('Order Amount')['Discount'].sum().sort\_values(ascending= False).head(30)

**Post - Analysis**

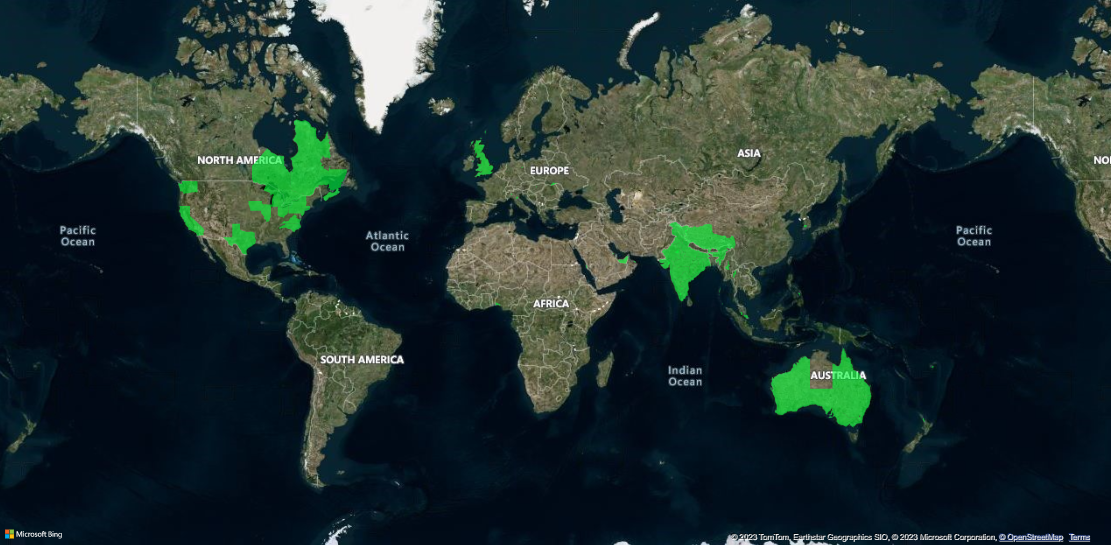
**By checking the shipping state, we can say that offline expansion should be in Maharastra.**

**And the Store should be in Mumbai because we have so many customers from Mumbai and next second is Pune which is very near to Mumbai and it can be useful for shipping purposes too.**

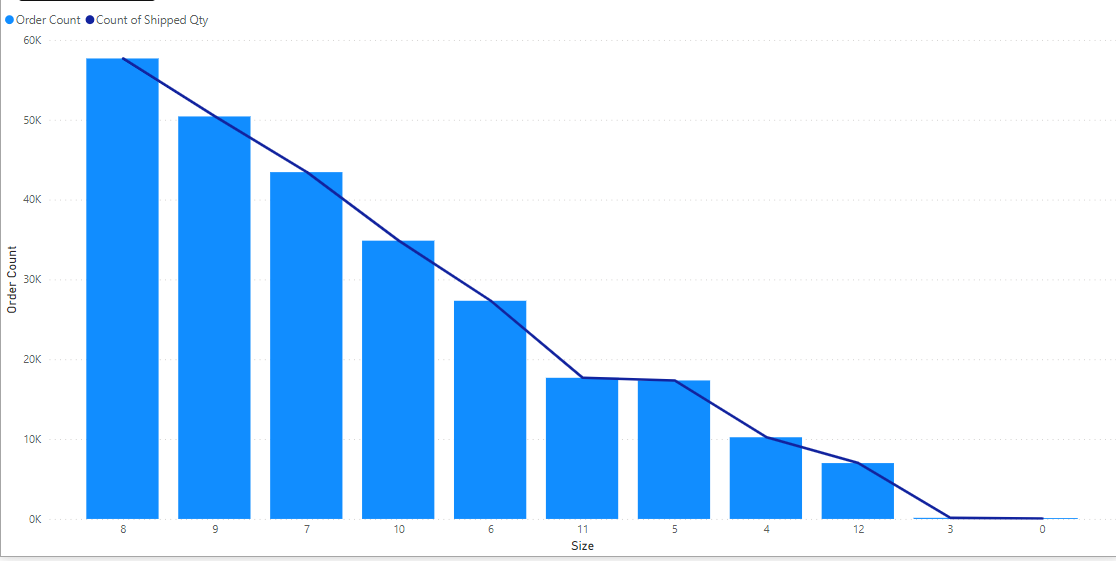
**As per Analysis, the variants which are sold more than others are size 8, size 9, and size 7. So, it is best to have a greater number of these types of variants to run the business smoothly and increase revenue.**

**As our sales are great in the second half of the year so it is better to increase our inventory in the last 6 months and plan according to regarding the variants of sizes 9,8,7 to make an impact on the business**

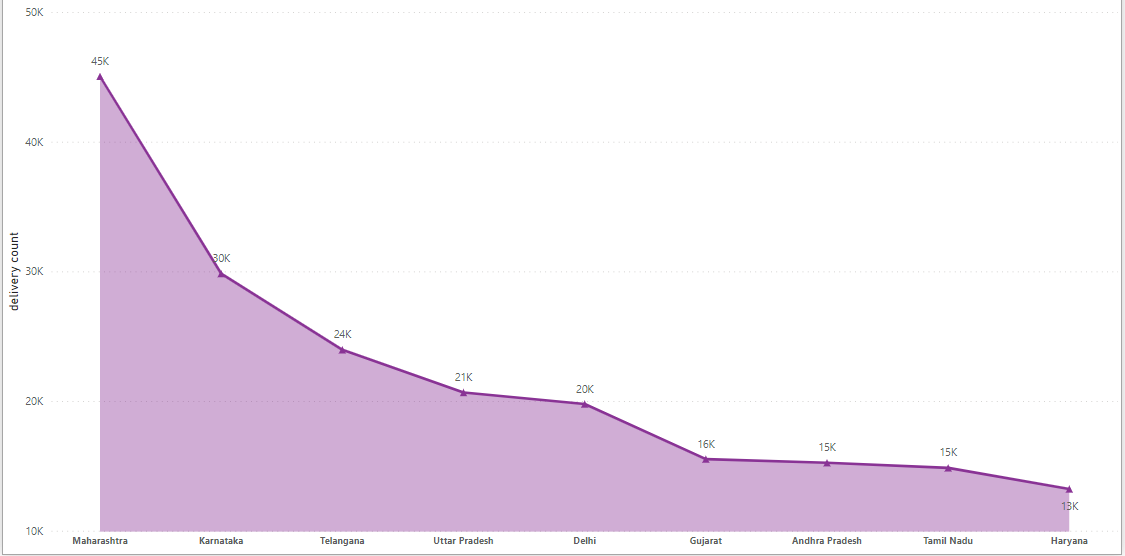
**From Analysis Flat Discount is not needed for the products. But a Partial/Upto discount is needed to boost the sales. Most Probably a discount is needed for the order amount which is more than 2500 and for those below the 2500 amount products we do not need discounts because we already have existing customers. Sometimes customers expect discounts for those products too, in that case, we can put discounts when there are any festivals and public holidays for boosting the brand.**



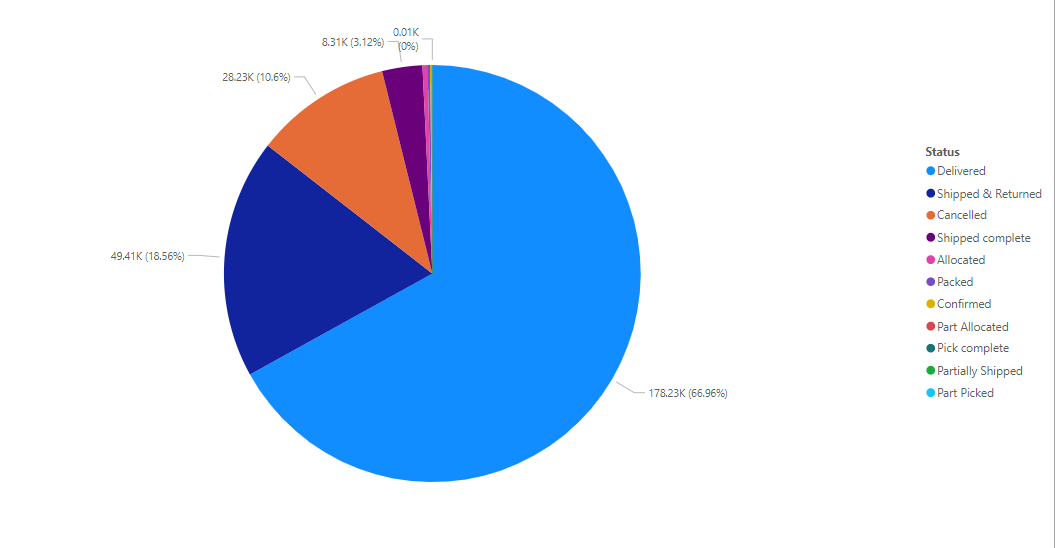
As per the above visual, the visual represents the Total shipping cities of the orders.



Order count and Shipped Quantity according to the variants of the products.



Top cities which have more delivered orders.



Status of the order wise distribution

Created a Matrix type visual for the Count of discounts within all variants of products according to order amounts.

