Capstone Project | BI Data Science Learning Path |

**Abstract**

**Project 2**

**Customer Segmentation for Retail**

In today’s fast moving world of marketing from product-orientation to customer-orientation, the management of customer treatment can be seen as a key to achieve revenue growth and profitability. Knowledge of customer behavior can help companies re-evaluate their strategies with the customers and plan to improve and expand their application of the most effective strategies. One of the big challenges in customer-based organizations is customer cognition, understanding the difference between them, and scoring them. Now with all capabilities we have, using new technologies like machine learning algorithm and data treatment we can create a very powerful framework that allow us to best understand customers need and behaviors, and act appropriately to satisfy their needs. In this project, machine learning’s various algorithms are applied to identify the hidden patterns in the data for better decision making. The concept of which customer segment to target is done using the customer segmentation process using the clustering technique. Here the clustering algorithm used is K-Means algorithm which is partitioning algorithm, to segment the customers according to the similar characteristics.

**Tools / Skills used**

1. Python Programming
2. Jupyter Notebook
3. Pandas
4. NumPy
5. Matplotlib
6. Seaborn
7. Exploratory Data Analysis
8. Data Visualization
9. Machine Learning
10. K-Means
11. RFM(Recency, Frequency, Monetary)

Introduction to the project

**Customer Segmentation for Retail**

In this project, we have worked on a model based on RFM model Recency, Frequency, Monetary and K-Mean algorithm to resolve the challenge of understanding the needs and behaviors of customers and to satisfy the needs of them. This model will allow us to use Clustering, Scoring and Distribution to have a clear idea about what action we should take to improve customer satisfaction.

**Implementation**

**Workflow:**

**Import Data**

EDA

Result

Result

Prediction

RFM

K-Means

Customer Segmentation

Data Preprocessing

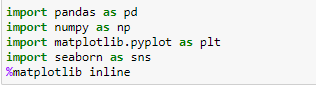
Data Visualization

**Modeling:**

1. **Customer Segmentation:** In this era, there is a strong competition in the business world, the organizations have to increase their profits and business by satisfying the demands of their customers and attract new customers according to their needs. The identification of customers and satisfying their demands is a complex task because customers may differ in demands, tastes, preferences and many more. This problem can be solved by Customer segmentation which clusters the customers into groups sharing the same properties or behavioral characteristics i.e. it make homogenous groups from data. The customer segmentation technique allows the business to make better use of marketing budgets, gain a competitive edge over their rival companies, demonstrating the better knowledge of the needs of customers.
2. **Clustering and K-Means Algorithm:** Clustering is the process of dividing a set of physical or abstract objects into groups of similar objects. The K-Means algorithm is one of the most popular clustering algorithms. Compared with other clustering algorithms, the K-Means algorithm is not only faster in calculation but it can also reduce the misclassification rate of data. The accuracy of this algorithm depends on initialization conditions and the number of clusters.
3. **RFM:** As a popular tool of customer value analysis, RFM has been widely used for measuring customer lifetime value and in customer segmentation and behavioral analysis RFM stands for **Recency** (Time interval between the customer’s last purchase date and the end date of a statistical period), **Frequency** (number of purchases made by the customer during the statistical period) and **Monetary** (total amount the customer spends in purchase during the statistical period). It provides an effective analysis for decision makers in order to target their customers and develop appropriate marketing strategies according to their previous behaviors. Although RFM model has been widely applied in various areas, its simplicity threatens its effectiveness since it does not consider the customers’ relationship and changes in customers’ behavior.

**Code Snippets:**

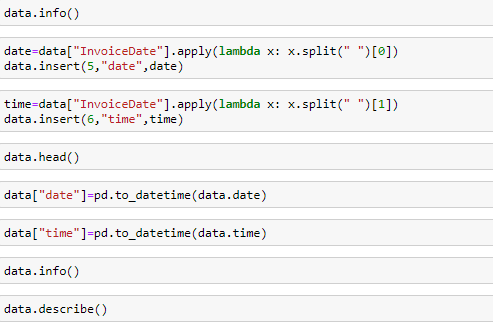
**Importing Libraries:**

****

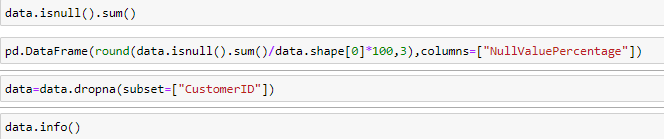
**Importing Data:**

****

**Understanding columns and their data types:**

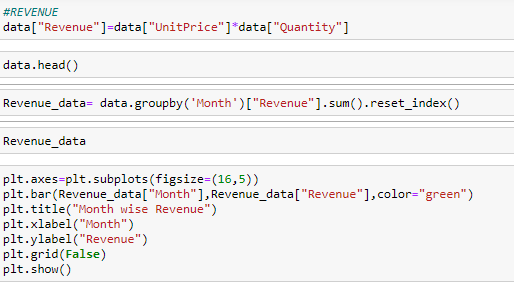
****

**Finding and dealing with Null Values:**

****

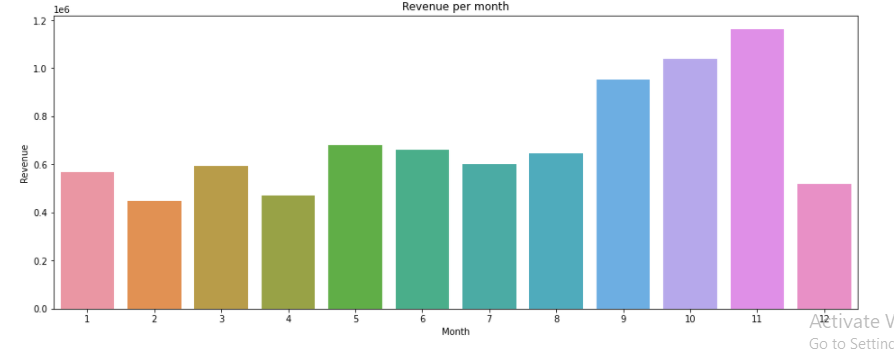
**Conclusion:** Approximately 24% of data from Customer ID is null. We can’t solve this problem by applying central tendencies so deleted the data where Customer ID is null.

**Revenue per Month:**

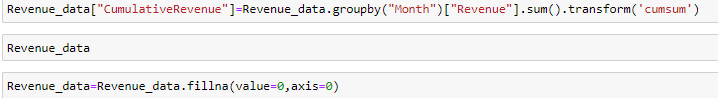
****

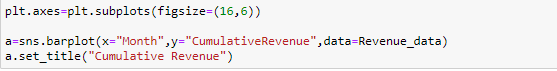
**Conclusion:** The highest boon in revenue was in the 9th month. Whereas the revenue for 11th month is the highest.

**Visualization snippets:**

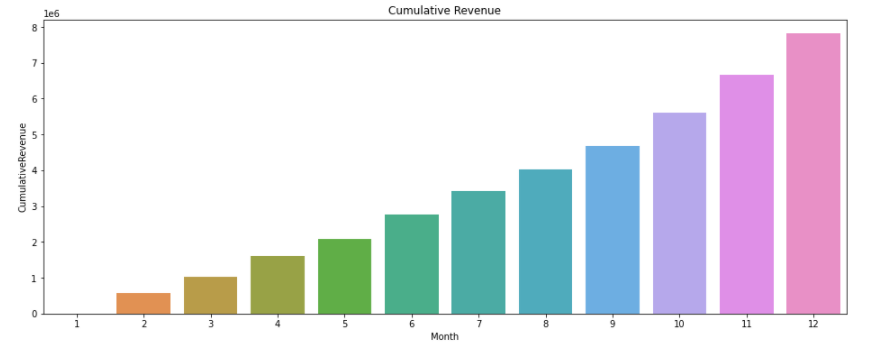


Cumulative revenue:

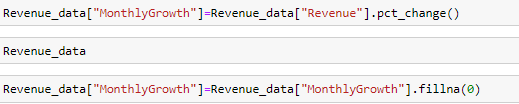




**Visualization snippets:**

****

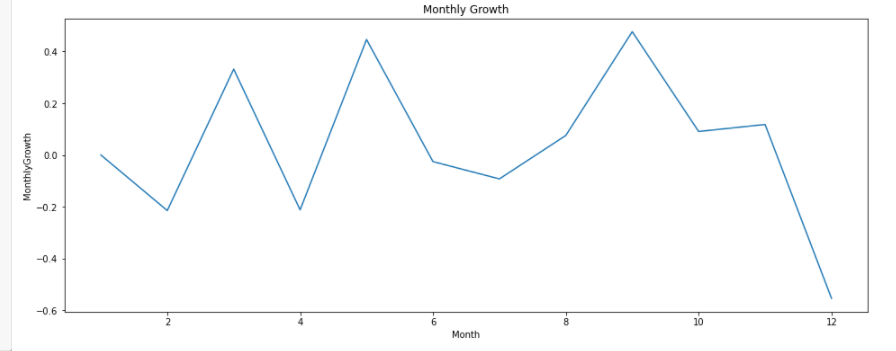
**Monthly change in Growth Rate:**

****

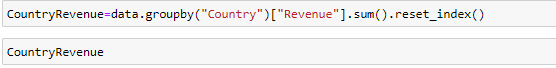
****

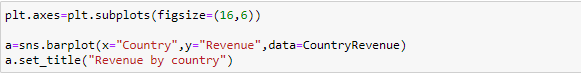
**Conclusion:** There were so many ups and down in growth rate , which shows that the growth rate for all the months are not increasing or decreasing at constant rate.

**Visualization snippets:**



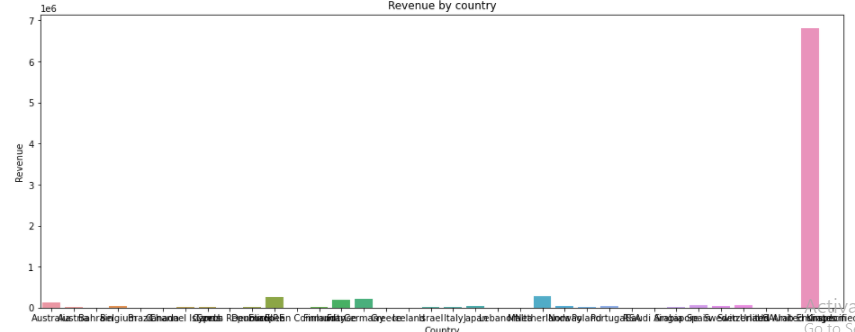
Revenue by Country:





Conclusion: Most of the Revenue is generated by United States. And there is a huge difference between the revenue generated by United States and other countries is very much.

Visualization Snippets:



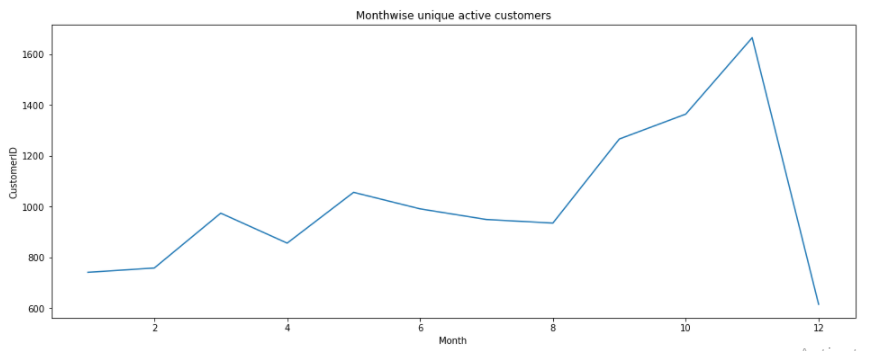
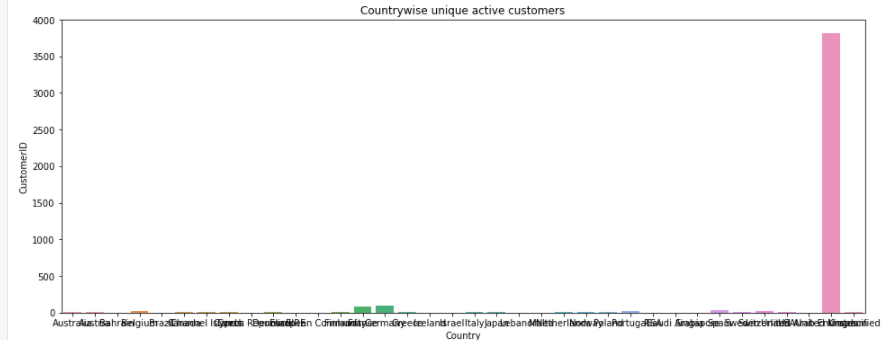
Active customers:

By month:

By Country:

Conclusion: There were continuous ups and downs in active customers but in between the period of 5th to 8th month, company has lost so many active customers but soon after this face there comes a complete boon for the company as the number of customers increased rapidly.

Visualization Snippets:

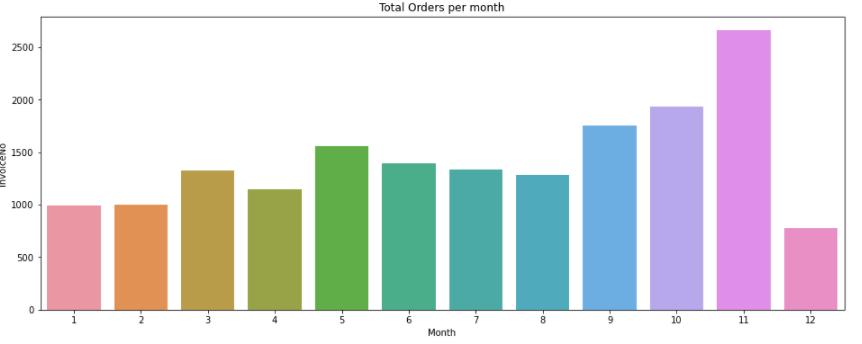
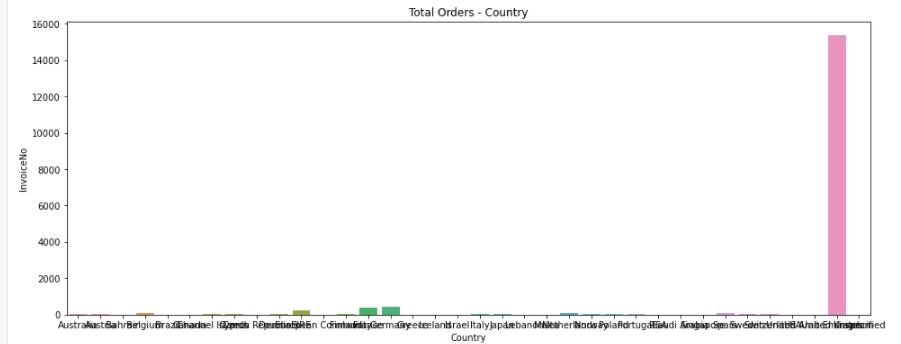
 

Total orders :

By Country:

By Month:

Visualization Snippet:

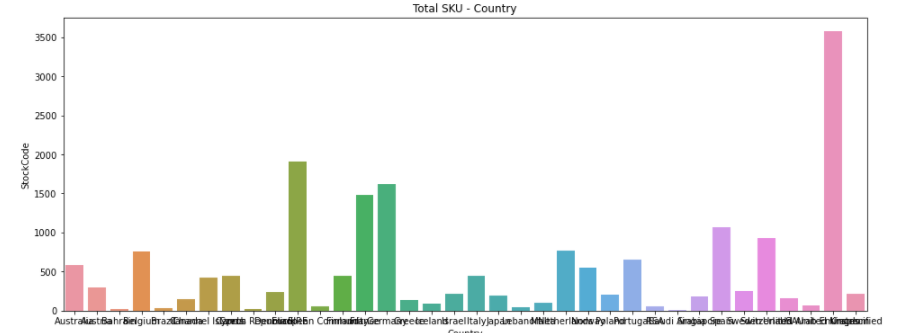
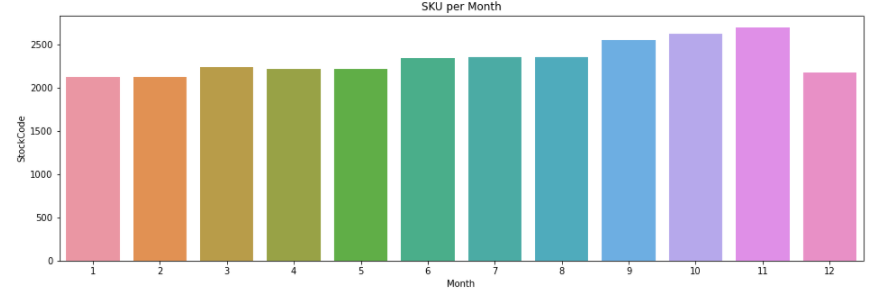
 

Distinct count of Stock Code:

By Country:

By Month:

**Visualization Snippets:**

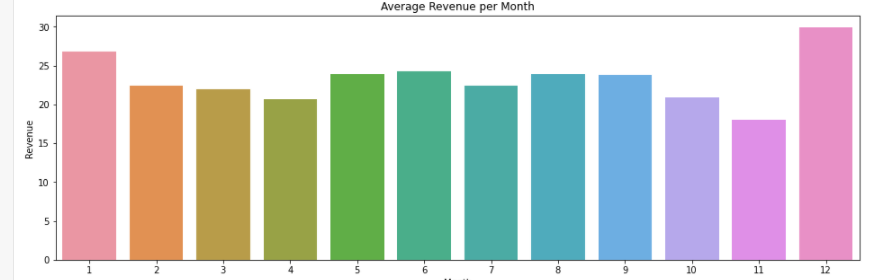
** **

**Average monthly revenue per order:**

****

****

**Visualization Snippets:**

****

**Conclusion:** The average revenue per order is highest for the 12th month. We are considering that we don’t have complete dataset of 12th month but the customers who visited the site in 12th month spent more on the products.

**New and Existing Customers:**







Monthly Total Customers by User Type:







Conclusion: The existing customers were highest for the 11th month but the new customers were highest for the1st month that means highest number of customers added with the company in 1st month.

Total Revenue per Month by User Type:



Conclusion: The revenue generated by the existing customers in the 11th month was highest. One reason could be because the total number of active customers in 11th month was also highest.

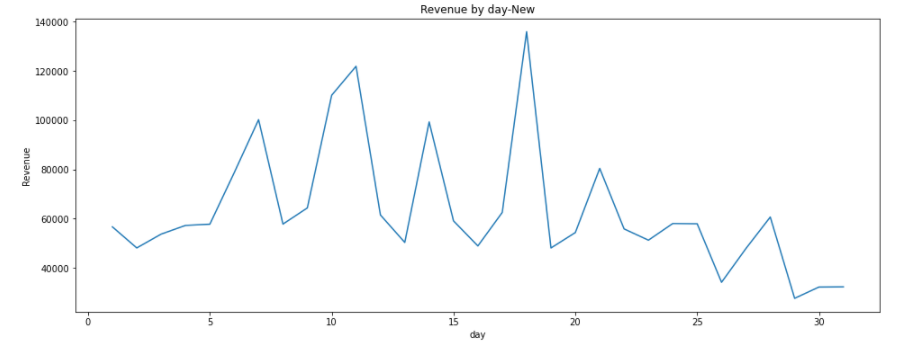
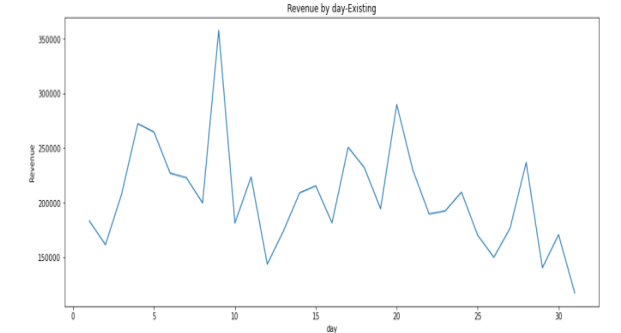
Total Revenue per month by day:







Visualization Snippets:

Monthly Retention Rate:





Customer Segmentation

(Using RFM Method)









RFM (Recency, Frequency, Monetary)

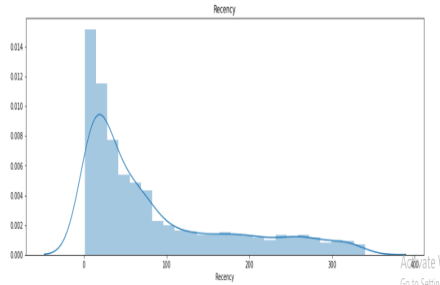
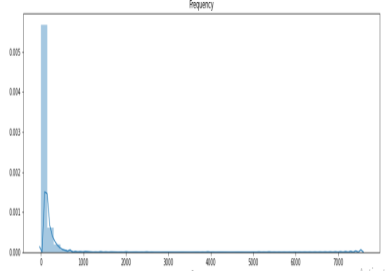
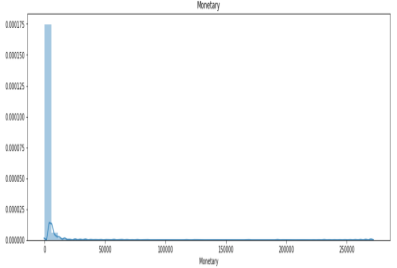






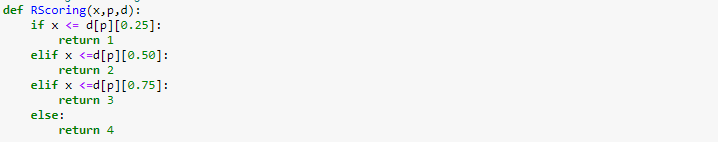


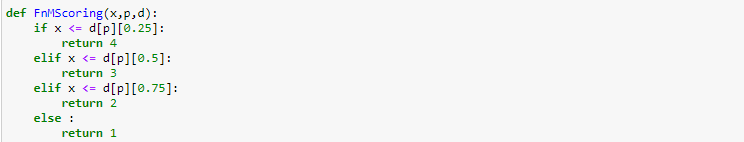
Visualization Snippets:

**Recency Frequency Monetary**

**Creating RFM Segments:**

****

****

****

****

****

**Analyzing Best, Medium and Worst Customers:**

**** 

**Conclusions:**

1. There are 1649 customers who do visit and purchase a good amount of products regularly.
2. There are 1336 customers who comes under the medium category i.e. they purchase the products more often.
3. Also there are 1259 customers who do not purchase company products very often or they have purchased products only once or twice.
4. So the number of regular customers is more in numbers than the number of customers who only purchased the company products once or twice.

**Future Scope:**

1. The company should focus more on the regular customers. The company should focus on the products customers are more interested in and should improve the quality of those products so the customers should remain satisfied.
2. There is also one more factor that can increase the amount of best customers for the company and that is publicity. Also they can announce some sort of discounts on some products so more and more customers can be attracted.
3. The company has 1336 medium category customers. To attract such customers the company should bring new policies and should also focus on customizing the product according to the customer’s demands so to attract these customers to purchase more and more products.
4. Company should get the information about the customers who do not purchased more company products and resolve the problem they are facing.