ABSTRACT

Study of customer behavior in online shopping usually deals with identification of customers and their buying behavior patterns. The aim of such studies is to make certain who buys where, what, when and how. The results of these studies are useful in the solution of marketing problems. Various studies on customer purchasing behaviors have been presented and used in real problems. For analysis of customer behaviors data mining techniques are consider more effective. The target of this paper is to analyze behavior of such people who are visiting the online shopping sites and spending their time there, surfing for different stuff. It would also be taken into account that how many people are there and how many of them are actually shopping. In this paper, different queries are applied to mine the database of a specified site which results in analysis of customer behavior towards online shopping.

1.4 Scope of Project

Customer Behaviour Analysis involves the prediction of Customer Behaviour with the help of previous statistics and data. This project is aimed at aiding the shop owners and retail markets to keep a constant tab on their customer base.

A customer analysis will do three main things:

* Identify the target customer
* Understand the needs of the customer
* Show how the company’s product or service meets the customers’ needs or wants

Our project uses Machine Learning algorithms like K-means algorithm which uses the technique of centroid clustering. We use such clustering algorithms to predict the future sales or the next visit of a particular customer. This in turn helps owners to know their customers better, thus leading to an increase in sales. Retail Markets can also provide regular customers with incentives and promotional offers which ultimately further increases the shop’s income.

1.5 Organization of Project

Our project is mainly organized into 5 categories namely:

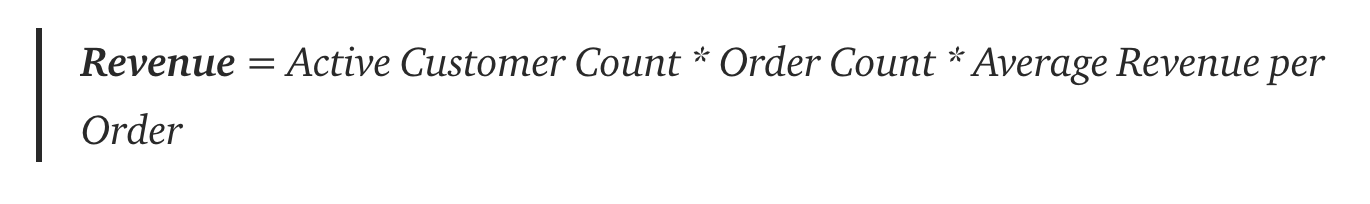
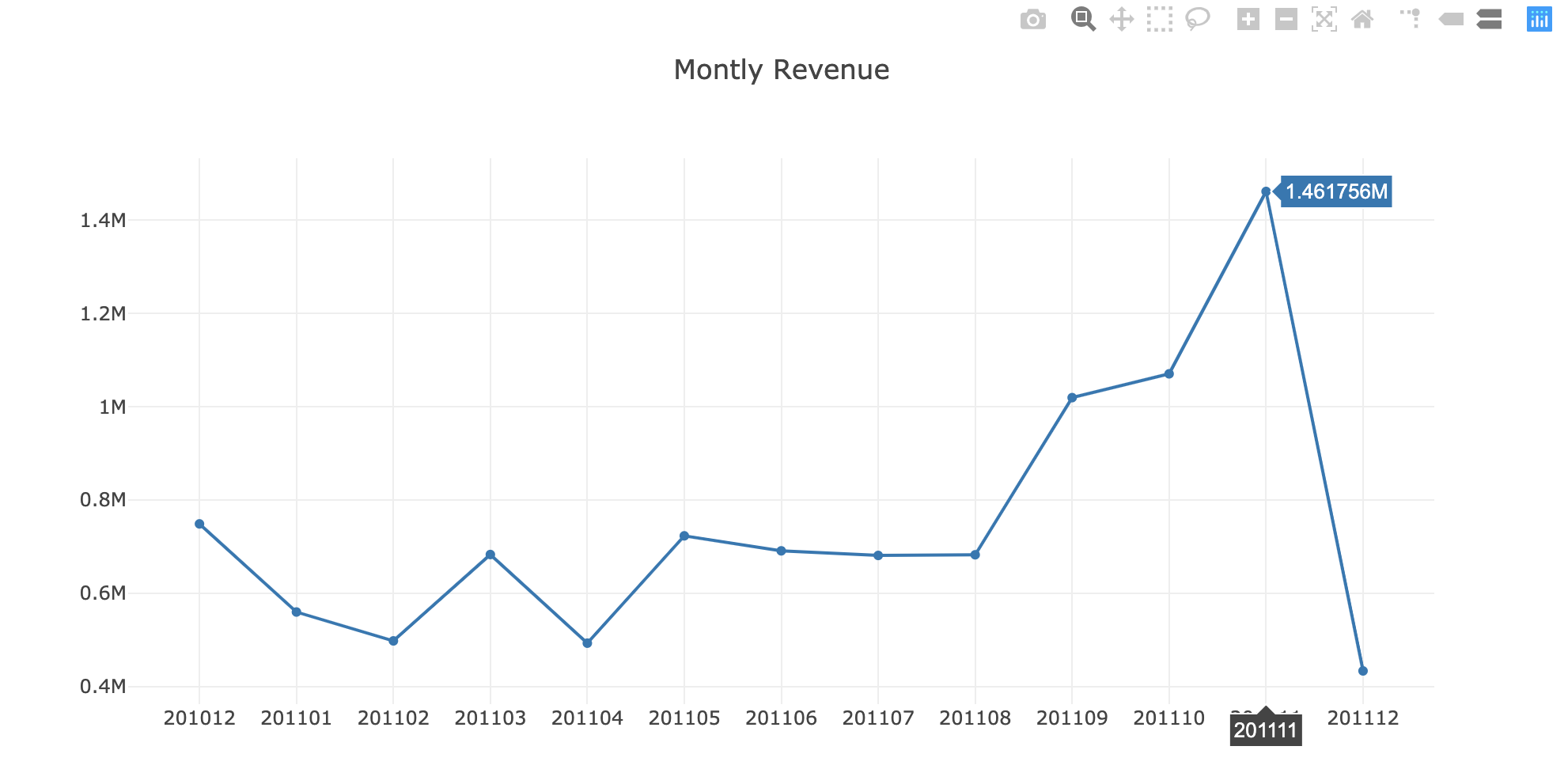
1. Know your Metrics
2. Customer Segmentation
3. Customer Lifetime Value Prediction
4. Predicting Next Day Purchase
5. Predicting Sales

Know Your Metrics

This phase involves a thorough understanding of the dataset and the variables present in it. Our dataset has above 5,00,000 entries and there are 8 columns in the dataset namely InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID and Country. Our time frame is from December 2010 to December 2011.

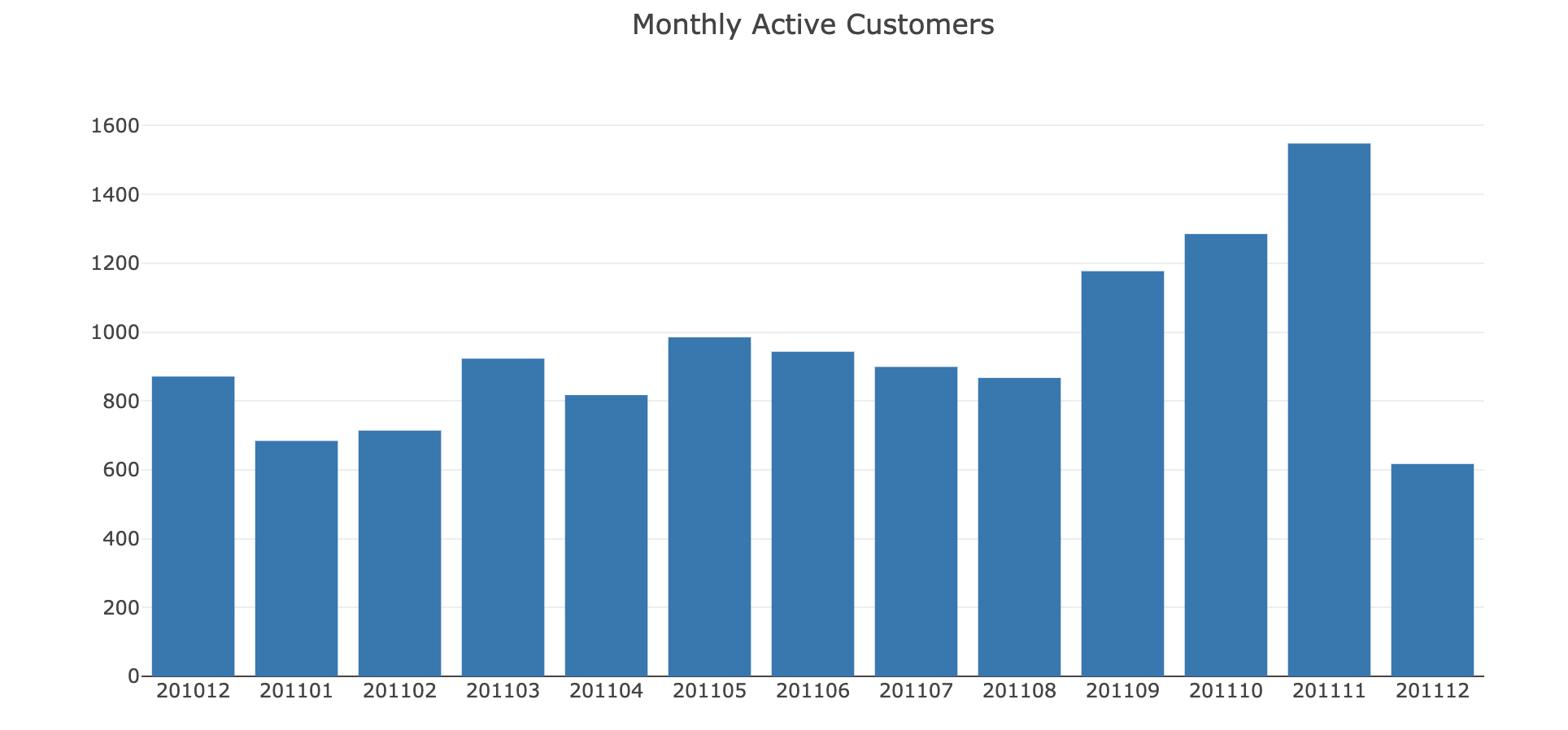
Revenue Calculations

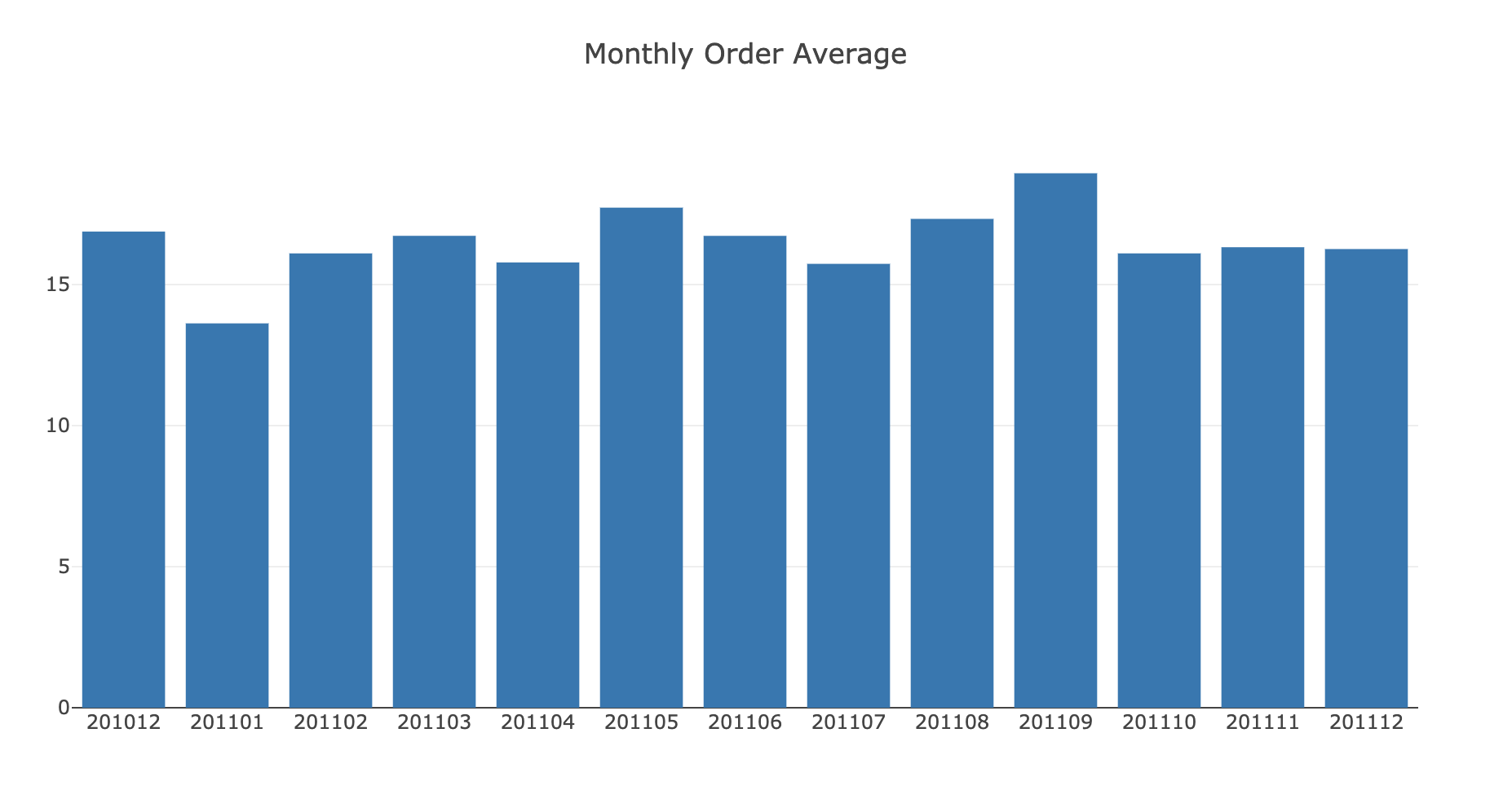
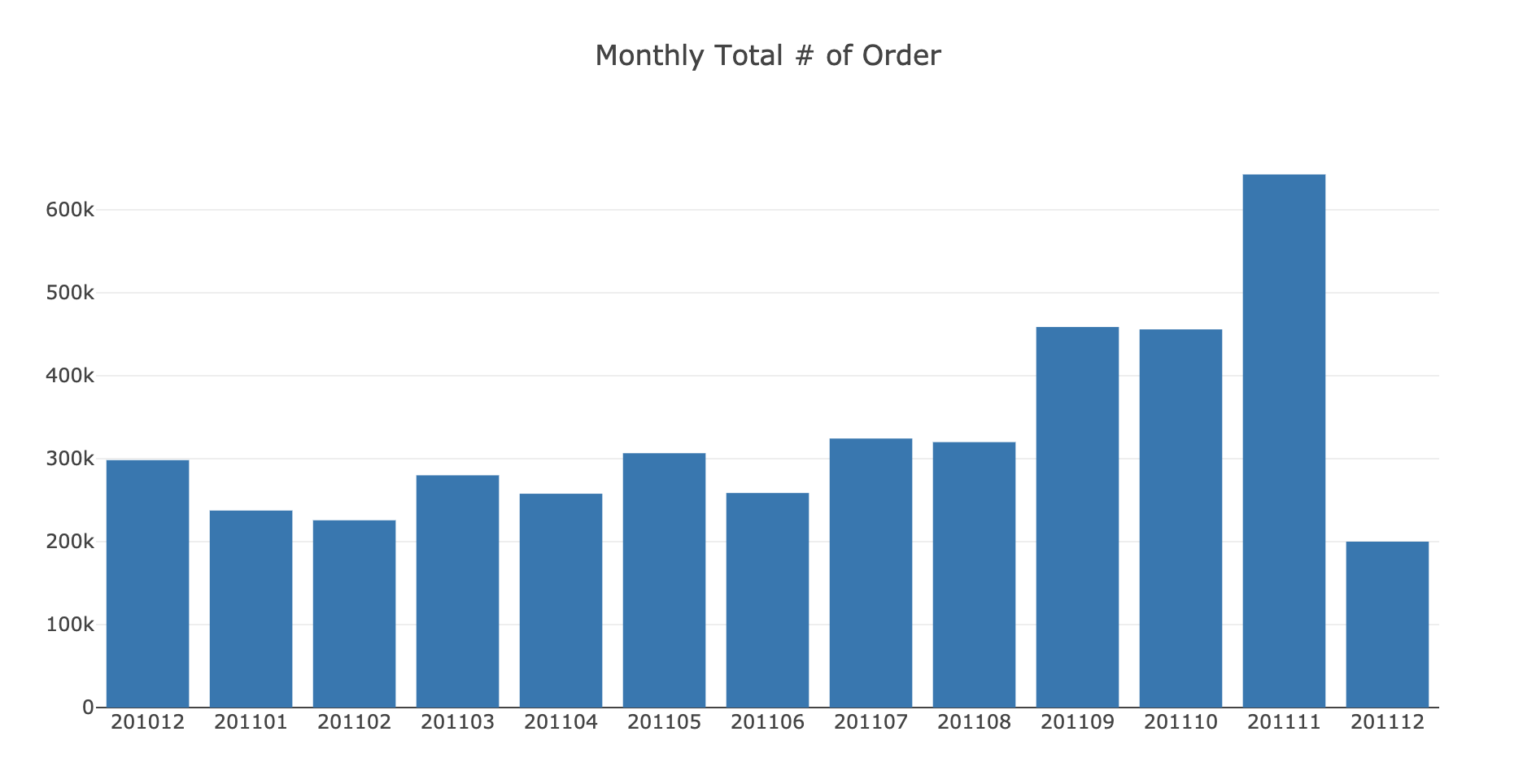
Initially we calculate the Revenue of a particular month by using the following formula:



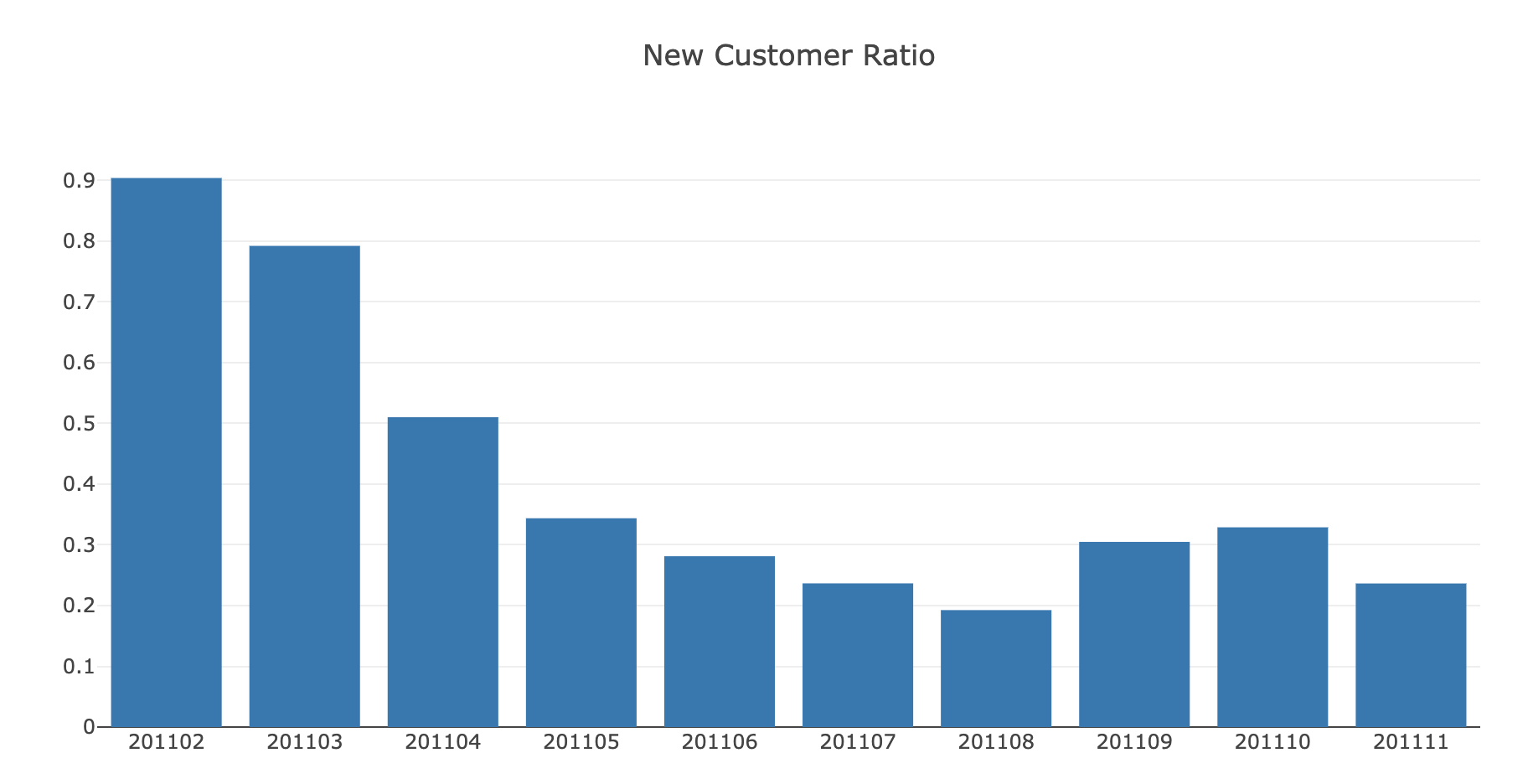
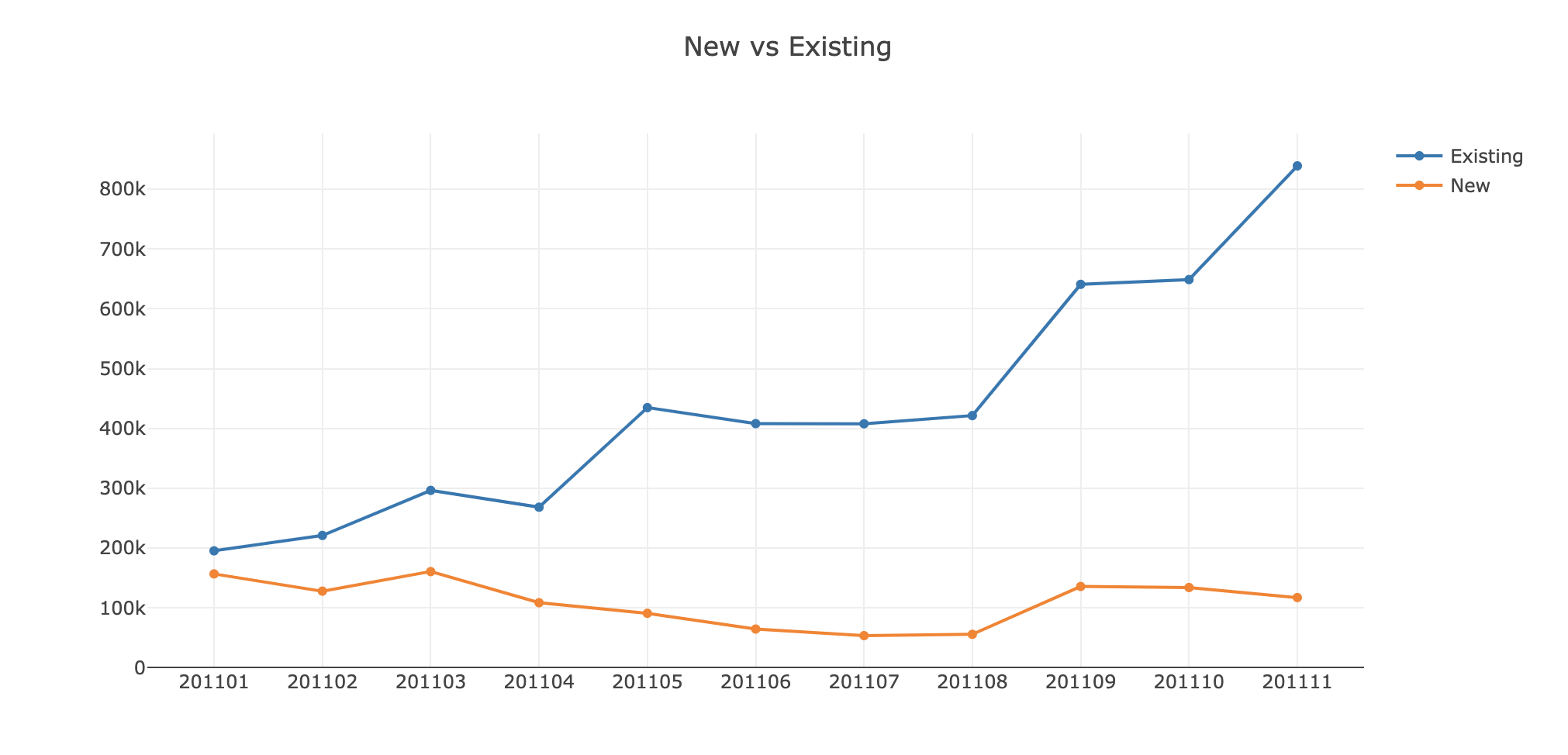
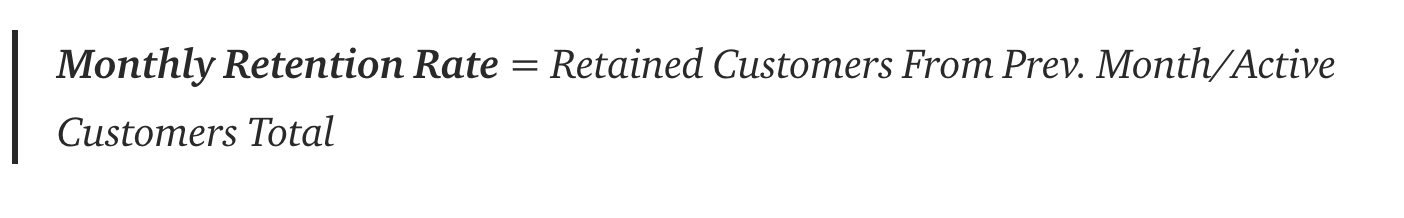
Monthly Active Customers , Monthly Order Average and Monthly Order Count

Next, we calculate the number of Monthly Active Customers and Monthly Order count for each month, followed by the Average Revenue per Order which will enable us to further analyze the Revenue pattern of the shop for each period i.e. for each month during the 1-year period from December 2010 to December 2011.



New Customer Ratio and Monthly Retention Rate

* A New Customer is defined as a person whoever did his/her first purchase in the time window we defined. By calculating the above, we can segregate the revenue into two classes namely **Revenue for Existing Customers** and **Revenue for New Customers.**
* Retention rate should be monitored very closely because it indicates how sticky is your service and how well your product fits the market. For making Monthly Retention Rate visualized, we need to calculate how many customers retained from previous month.

****