Lab Assignment

Date of Submission-19-09-2020

Question 1:

A firm has been providing services to its customers. The dataset **(complaint.csv)** contain following attributes

1. Ticket #: Ticket number assigned to each complaint
2. Customer Complaint: Description of complaint
3. Date: Date of complaint
4. Time: Time of complaint
5. Received Via: Mode of communication of the complaint
6. City: Customer city
7. State: Customer state
8. Zipcode: Customer zip
9. Status: Status of complaint
10. Filing on behalf of someone

Analysis Task

To perform these tasks, you can use any of the different Python libraries such as NumPy, Pandas, scikit-learn, and matplotlib.

[1] Produce the trend chart about the monthly registered complaints

[2] Generate a tabular output with frequencies of complaints

[3] Find which complaint types are maximum i.e., internet, network issues, etc.

[4] Provide the state wise status of complaint in the form of stacked bar

[5] Report which state has registered maximum and minimum complaint

[6] Which state has the highest and lowest percentage of unresolved complaints

Question 2:

Mart would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. An ML algorithm will predict demand accurately and factors like economic conditions including CPI, Unemployment Index, etc. Store runs promotional events throughout the year (which are the Super Bowl, Labour Day, Thanksgiving, and Christmas). Part of the assignment is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data.

Dataset Description (Name of Dataset:Mart.csv)

This is the historical data that covers sales from 2010-02-05 to 2012-11-01. Within this file you will find the following fields:

Store - the store number

Date - the week of sales

Weekly_Sales - sales for the given store

Holiday_Flag - whether the week is a special holiday week 1 – Holiday week 0 – Non-holiday week

Temperature - Temperature on the day of sale

Fuel_Price - Cost of fuel in the region

CPI – Prevailing consumer price index

Unemployment - Prevailing unemployment rate

Holiday Events

Event 1: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13

Event 2: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13

Event 3: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13

Event 4: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

Basic Statistics tasks

1. Which store has maximum sales
2. Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation
3. Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together
4. Provide a monthly and 6 month view of sales in units and give insights

Statistical Model

For store 2 develop Linear Regression model– Utilize variables like date and restructure dates as 1 for 5 Feb 2010 (starting from the earliest date in order). Hypothesize if CPI, unemployment, and fuel price have any impact on sales.

Learning Resources for Assignment

[1] Pandas

https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html

[2] Numpy

https://numpy.org/devdocs/user/quickstart.html

https://www.w3schools.com/python/numpy_intro.asp

[3] Python

https://docs.python.org/3/tutorial/