Comcast Telecom Consumer Complaints

Table of Contents

Introduction

This is the project report for the Project 3: Comcast Telecom Consumer Complaint" present as part to the "Data Science with Python" course of Simplilearn.

- Project submitted by : Krishnaveni Rajan
- Email Id: krishnaveni07.rajan@gmail.com
- Submission Date: 03-May-2021
- Language Used : Python

Problem Statement

Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a $2.3 million, after receiving over 1000 consumer complaints.

The existing database will serve as a repository of public customer complaints filed against Comcast. Using the dataset, help to pin down what is wrong with Comcast's customer service with the following information

1. Provide the trend chart for the number of complaints at monthly and daily granularity levels.
2. Provide a table with the frequency of complaint types.
3. Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
4. Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on "Which state has the maximum complaints?"
5. Which state has the highest percentage of unresolved complaints
6. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

Overview

The analysis request by the Comcast Telecom is to provide the insights of the complaints based on various factor with the data provided for the year of 2015. This is requires the "**Descriptive Analytics**" to be done to analyse the data from the historic data and

provide the insights about what has happened in the past. This information of Descriptive Analytics could be used the company to come-up with the corrective actions, improvements required and identify the gap in the existing process.

Phases of Project Work

The project work has been done using the following phases

1. Understanding the Problem statement
2. Review the Dataset to understand the data provided
3. Identify the columns in the Dataset that requires Data Wrangling
4. Break the analysis required for the solution into smaller chunks
    a. Come up with the step-by-step activity to be done (without starting the coding)
5. Write the Python code
    a. Execute the code and get results
    b. Ensure there is no errors
6. Complete the project and submit for Grading.

Data Dictionary

Data dictionary of the Comcast Telecom Complaints data.csv is as follows

- Ticket #: Ticket number assigned to each complaint
- Customer Complaint: Description of complaint
- Date: Date of complaint
- Time: Time of complaint
- Received Via: Mode of communication of the complaint
- City: Customer city
- State: Customer state
- Zipcode: Customer zip
- Status: Status of complaint
- Filing on behalf of someone : Confirm if the ticket is filled by the customer directly or being filed on on-behalf of others

Data Wrangling Performed

1. Date Format

Dates, Months and Quarter column were newly added to the Data frame to generated using the Pandas, “to_datetime”. This will help in generating the Daily & Monthly trend of complaints and also for getting Q3 insights.

2. Grouping of Data

Data set provided was grouped using ‘groupby’ function for getting

a. Date (Daily range)
b. Month (Monthly Trend)
c. Quarter
d. City (State wise)
e. Status wise
f. By the complaints received source (Received Via)

3. Status Based Changes

Values in the Status column was worked on to retain only Open and Closed status using the apply, lambda and if loop.

Functions and Libraries Used

Different functions and libraries were used in the R code while coming up with the program to deliver the required results

Libraries Used

a. numpy as np
b. pandas as pd
c. matplotlib.pyplot as plt
d. rlcompleter, readline
e. import sys

Functions Used

a. read_csv
b. head
c. to_datetime
d. groupby

e. count
f. show
g. value_counts
h. to_frame
i. reset_index

Conclusion

With the data analysis done on the dataset the 'Descriptive Analytics' inference is as follows

1. Total tickets raised in the year 2015 was 2224. In the year of 2015
2. On 6th Date there was maximum number of complaints in the Daily granularity.
3. June motnh had the maximum number of complaints in the Monthly granularity.
4. The Top 5 category of the complaint type and their frequency was as follows
    a. Comcast 83
    b. Comcast Internet 18
    c. Comcast Data Cap 17
    d. comcast 13
    e. Data Caps 11
5. In-sights for Q3 of 2015
    a. **Utah** had the maximum number tickets in "Open" Status which counted to 1.
    b. Out of the total 2224 tickets raised, "Resolved %" of tickets in Q3 was as follows
        i. Customer Care was 56.41%
        ii. Internet was 43.59%
6. In-sights for the year 2015
    a. **Georgia** had the maximum number tickets in "Open" Status which counted to 80.
    b. Out of the total 2224 tickets raised, "Resolved %" of tickets in Q3 was as follows
        i. Customer Care was 50.62%
        ii. Internet was 49.38%

NOTE: The Source code and Snapshots of the result are shared in separate documents.