Comcast Telecom Consumer Complaints Analysis

Description:

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. It will help to pin down what is wrong with Comcast's customer service.

Data Dictionary

- 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
- Zip Code: Customer zip
- Status: Status of complaint
- Filing on behalf of someone

METHODOLOGY:

Import Comcast telecom complaint data

- Import Data from relevant path
- Before importing the data Import different Importing packages and libraries for python
- like NumPy, SciPy, Pandas, scikit-learn, matplotlib

Basic data exploratory analysis

- Explore data
- Understanding the Dataset
- To gain insights from data we must look into each aspect of it very carefully. We will start with observing few rows and columns of data both from the starting and from the end
- Print the number of rows and columns of the Data Frame
- Print the names of all columns.
- Print the last n rows of the Data Frame
- Give Index, Datatype and Memory information
- Pre-processing
- Dealing with missing values
- Dropping/Replacing missing entries
- Replacing missing values with median values.
- Dropping the column as it has too many null values.
- Replacing 0 values with median values.

- Post Pandas Profiling
- Now we have pre-processed the data, now the dataset does not contain missing values, we have also introduced new feature. So, the pandas profiling report which we have generated after pre-processing will give us more beneficial insights.

- **Installing packages:**

  Commonly used packages are:

    - seaborn
    - matplotlib
    - bokeh
    - plotly

- Provide the trend chart for the number of complaints at monthly and daily granularity levels
    I. Lookout complaints over date and month
- Look out the status of complaints over the months
- Identify number of complaints state wise and plot using bar graph
    I. Which state has the maximum complaints
- Find top 10 complaints using counter and plot using bar graph
    I. Provide a table with the frequency of complaint types
    II. Which complaint types are maximum i.e., around internet, network issues, or across any other domains
- Calculating state wise complaint resolving rate
    I. Create a new categorical variable with value as Open and Closed
    II. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed
- Calculate percentage of complaints resolved till date - Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls

  Observations:
- The given data table has 2224 rows and 11 columns
- Data Types: `dtypes: int64(1), object(10)`
- Memory Usage: `memory usage: 191.2+ KB`
- There are no missing values in our dataset
- Customer complaints were highest in the month of June and on 24-06-2015
- Closed Cases: Highest in April and lowest in May
- Open Cases: Highest in June and lowest in April
- Solved Cases: Highest in June and lowest in April
- Pending Cases: Highest in June and lowest in April
- Georgia state has the highest number of complaints with a count of 288 cases followed by Florida h aving 240 cases and California having 220 cases
- Comcast has the highest frequency with count equal to 83, followed by Comcast internet with a co unt of 18 and Comcast data cap with a count of 17
- Georgia has the highest number of open and closed cases with count equivalent to 80 and 208 resp ectively