1. **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

**Analysis Task:**

* Import data into SAS environment.
* Provide the trend chart for the number of complaints at monthly and daily granularity levels.
* Provide a table with the frequency of complaint types.
* Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
* Create a new categorical variable with value as **Open** and **Closed**. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.
* 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
* Which state has the highest percentage of unresolved complaints
* Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

1. **DESCRIPTION : Directing customers to the subscription through app behavior analysis**

This project includes data for a website which has its main product in the form of subscription eg youtube red, or pandora premium. We have data about various customers such as their first\_open date for the app, their time, age, screen names, no. of screens, enrolled date, etc. There are three files here of total size 22MB. This is an artificial data but uncleaned so will require cleaning just like any real data. The data is for the first 24 hours because after that the free trial expires. App behavior is characterized as a list of app screens the users opened, the time they spent on those, the clicks on those pages, etc.

**OBJECTIVE**: The main task here will be to understand customer behavior and classify if he will **not** subscribe to the product. Once identified this customer will be targeted with promotions to convert them and hence increase subscription levels for premium services.

(Data uploaded)

**Techniques used**: Plotting, data manipulation, Classification models (Logistic regression, regularization), K fold cross-validation, Grid search, and Feature selection algorithms.

1. **DESCRIPTION:** The global fashion industry is valued at three trillion dollars and accounts for 2 percent of the world's

GDP. The fashion industry is undergoing a dramatic transformation by adopting new computer vision and machine learning and deep learning techniques. Here, we'll look at a hypothetical situation.

**OBJECTIVE**: build a virtual stylist assistant that looks at customer Instagram and Facebook images and classifies what fashion category they are wearing either bags dresses and pants. The virtual assistant should help the retailer detect and forecast fashion trends and launch targeted marketing campaigns. Here, data set contains images of bags, shoes and dresses. And we're asking the deep network to classify the images into 10 classes. Once that happens we can take a user on instagram and have targeted campaign for him/her based on what category of clothes he/ she loves to wear

Data set is of 148MB. Its rows are 0s and 1s, signifying the greyscale for images. Each row represents a thing like a bag, shoes, dress, shorts etc.

70k images

60k training dataset

10k test dataset

These are 28X28 pixel grayscale images

(Data uploaded)

Example : <https://www.youtube.com/watch?v=9X_fP4pPWPw>

1. **DESCRIPTION:** The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numeric input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

**OBJECTIVE:** The aim of this project is to build a classifier that can detect credit card fraudulent transactions. We will use a variety of [*machine learning algorithms*](https://data-flair.training/blogs/machine-learning-algorithm/) that will be able to discern fraudulent from non-fraudulent ones. By the end of this machine learning project, you will learn how to implement machine learning algorithms to perform classification.

Link: [Data Science Project - Detect Credit Card Fraud with Machine Learning in R](https://data-flair.training/blogs/data-science-machine-learning-project-credit-card-fraud-detection/)

Data Size: 144MB