Consumer Complaints in Banking Sector.

Abstract:

This Use Case focuses on exploring and analyzing Consumer Finance Complaints data, to find how many similar complaints are there in relation to the same bank or service or product. These datasets fall under the complaints of Credit reporting, Mortgage, Debt Collection, Consumer Loan and Banking Accounting. By using data mining techniques, cluster analysis, AWS Services as well as predictive modelling is applied to obtain valuable information about complaints in certain regions of the Country. The banks that are receiving customer complaints filed against them will analyse the complaint data to provide results on where the most complaints are being filed, what products/ services are producing the most complaints and other useful data. Our model will assist banks in identifying the location and types of errors for resolution, leading to increased customer satisfaction to drive revenue and profitability.

Technologies Used:

1. Amazon S3 Bucket. - Storing Raw data
2. Spark – ETL processing (PySpark/Spark Scala/Spark SQL)
3. Amazon Redshift – Storing processed Data

Problem Statements:

1. Clean and Transform data for proper processing and getting complete insights without any garbage values
2. Transferring and Storing data in Cloud.
3. Find the number of complaints for which the Company response is currently "in progress".
4. Which company has the maximum consumer complaints.
5. Which Companies is able to solve issues of customers (on the terms of Company response and timely response)
6. Which companies are having least response time for a complaint raised?
7. Find the issue that occurred mostly.
8. Which are the Top 5 states having the highest volume of complaints coming.
9. Which are the Top 5 companies people complaining the most.
10. Which product has the most number of complaints.

Indian Agriculture Analysis

Abstraction:

This project is used to extract information from agriculture data and to give suggestions regarding crops and make future decisions so that agriculture can be carried out in a planned manner.

Problem Statements:

1. Finding trends in crops in terms of production, area.

2. State wise suicide rate of the farmers

3. Which year has Highest cultivation of Pulses

4.Which year has Highest cultivation among rice, wheat and coffee.

5. Which state yields highest crop per year?

6. Finding how different factors like temperature affect production

7. Year wise suicide rates of farmers

8. Which state receives highest rainfall in which year and what is the annual rainfall?

9. Crop wise cost of cultivation versus cost of production

10. Highest crop production in particular season.

Technology

1. Amazon S3 Bucket- store the raw data
2. Spark- for ETL transformation and visualization(PySpark/Spark Scala/Spark SQL)
3. Amazon Redshift- for storing the processed data

Covid Data Analysis

Abstract :

The project is to perform data analysis on Covid data across the globe and extract the meaningful information from the dataset which will help in taking quick and informed decision so that will help in preventing further spread of Covid virus.

Problem Statement :

1.Clean and transform data for processing

2. ETL operations on dataset

3. Storing the modified data in Azure

4.Display total cases ,new cases ,recovered cases and deaths.

5.Which country in Distinct WHO region highest cases in till date.

6. Total no of confirmed cases over between a certain date.

7.Create a geographic Map Based on total cases .

8.Date of first confirmed case in a particular region.

9.Date on which max no. cases were reported in a country.

10.Line chart showing total cases, deaths &recoveries of a particular country .

11.Bar Chart showing cases, death and recoveries reported per day/week reported.

12.Display the highest recovery rate of each month .

13.No. of active cases vs critical cases of a country.

14.Average active cases vs recoveries recorded in a day .

Technology

1. Amazon S3 Bucket- store the raw data
2. Spark- for ETL transformation and visualization(PySpark/Spark Scala/Spark SQL)
3. Amazon Redshift- for storing the processed data