# Declaration

We **ABDUR REHMAN (BCSF16E031), ARSALAN ALI (BCSF16E060), HASSAN WALI (BCSF16E010)** hereby declare that we have produced the work presented in this thesis, during the specified period of study. We also declare that we have not taken any material from anywhere except referred to wherever due that amount of plagiarism is within acceptable range. If a violation of HEC rules on research has occurred in this thesis, I shall be liable to punishable action under the plagiarism rules of the HEC.

Date: _____ Signature of the student: _____

**ABDUR REHMAN**

**BCSF16E031**

**ARSALAN ALI**

**BCSF16E060**

**HASSANWALI**

**BCSF16E010**

# DEDICATION

The whole project is dedicated to our parents who are our source of inspiration and gave us strength of not giving up.

Secondly, we dedicate to Allah Almighty who gave us hard working power and skills.

# ACKNOWLEDGEMENTS

we would like to acknowledge everyone who prayed for us and played a role in accomplishing this project successfully. Without you we would never had reached this success. Secondly, we would acknowledge our parents who supported us with love and affection.

we would acknowledge teachers specially Sir Fahad Maqbool who provided us guideline throughout our work. Thank you all for your kind support.

**ABDUR REHMAN**

**BCSF16E031**

**ARSALAN ALI**

**BCSF16E060**

**HASSAN WALI**

**BCSF16E010**

# ABSTRACT

## Time Series Sale Forecasting

Time Series Analysis is a method for examining time series data in order to extract significant statistics including forecasting, monitoring or even feedback and other characteristics of the data. It plays an important role in many fields such as stock market, business, exchange, weather, electricity demand, usage and cost of products like gasoline, electricity etc.

Time Series Analysis for predicting future values has been a major topic since a few years ago. Historically, Time Series Analysis has been around for centuries and its evidence can be seen in the field of astronomy, it was used to study movements of planets and the sun in old times. But now in modern times, it is used in practically every field, from day to day business issues to complicated scientific research.

Time Series Analysis has different types, such as Descriptive analysis in which we determine trend or pattern in a time series using graphs and other tools, Spectral analysis aims to separate periodic or cyclical component in time series, Forecasting referred to business sales prediction and stock market and Intervention analysis to determine if an event can lead to change in time series.

Time Series Forecasting is an important part of Machine Learning. It uses different technologies such as Artificial Neural Networks, Machine Learning, support vector machines and Markov models etc. The important advantage of Time Series Analysis is that it can be used to understand past as well as predict future. Thus sales forecasting can be very useful for a Business. Firms or Companies focus on sales for good reason. Companies should grow their sales in order to exist in the market. Therefore, it becomes a point of prime importance for a business to forecast where it thinks sales are heading in the future, this will not only help a company in managing and tracking sales but also help in making important decisions at the right time.

In our final year project, we similarly perform a Time Series Analysis on a Business Data by using Artificial Intelligence models and predict Sales in the future.

# TIME SERIES SALE FORECASTING

*By*

Abdur Rehman

Arsalan Ali

Hassan Wali

A Documentation Submitted to Department of CS&IT

University of Sargodha

In partial fulfillment of the requirement for the degree of

BS (Computer Science)

# Contents

# Project Proposal

## Project Title

Project name is Time Series Sale Forecasting. Sales Forecasting will predict future values of sales using appropriate dataset. Time series use also in non-stationary data like economic, weather, stock price, and retail price as well in sales.

## Project Team

### Project Manager:

- DR. FAHAD MAQBOOL

### Project Members:

- ABDUR REHMAN (BCSF16E031)
- HASSAN WALI (BCSF16E010)
- ARSALAN ALI (BCSF16E060)

## Problem statement:

Many time series problems have contagious observations, such as one observation each hour, day, month or year. A time series where the observations are not uniform over time may be described as discontigous. In sales forecasting we observe the past values to predict the future sales as per criteria. This analysis also decreased the problem using ARIMA MODEL but still there were some technical issues, so we are using PROPHET MODEL to overcome these all problem.

## Executive Summary:

Time Series Analysis is a method for examining time series data in order to extract significant statistics including forecasting, monitoring or even feedback and other characteristics of the data. It plays an important role in many fields such as stock market, business, exchange, weather, electricity demand, usage and cost of products like gasoline, electricity etc..

Time Series Analysis has different types, such as Descriptive analysis in which we determine trend or pattern in a time series using graphs and other tools, Spectral analysis aims to separate periodic or cyclical component in time series, Forecasting referred to business sales prediction and stock market and Intervention analysis to determine if an event can lead to change in time series.

## Objective

- Data Filtration and Pre processing

- Visualizing sales time series
- Trend, Seasonality, Noise
- ARIMA and SARIMA Model implementation
- Producing and visualizing forecast and displaying patterns using Prophet Model
- UI using Flask framework of python

## Motivation

We are using new and fully featured Model called prophet model which is new technology towards time series analysis. It is best for predicting the future sales for business and policy planning, our analysis will be very interesting because it will predict the future sales using past values.

## Requirements

- Package Manager Anaconda
- IDE Jupyter Notebook
- Different Libraries Numbpy, Matplotlib, Pandas etc.

## Features of Project

- Login
- Signup
- Exploratory data analysis
- Trend and seasonality Visualization
- Specific Model Prophet
- Cyclic variations
- Graphical User Interface

## Implementation Tools and Techniques

### Programming Language:

- Python 3.8
- HTML and CSS for user interface
- MYSQL
- ARIMA MODEL and PROPHET MODEL

## References:

Emanuela Furfaro. (2018). *Data Analysis*. Retrieved from GitHub: https://emanuelaf.github.io/pred-sales.html

GURCHETAN SINGH. (2018, February 8). *Bootcamp*. Retrieved from Analytics Vidhya: 2018

*Introduction to Flask*. (n.d.). Retrieved from Pymbook: https://pymbook.readthedocs.io/en/latest/flask.html

Jason Brownlee. (2019, August 21). *Topics*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/time-series-forecasting/

Li Susan. (2019, January 15). *Time Series Analysis*. Retrieved from Towards Data Science: https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b

Marco Peixeiro. (2019, August 7). *Visualization*. Retrieved from Towards Data Science: https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775

*Projects*. (2019). Retrieved from Forecasting at Scale: https://facebook.github.io/prophet/

Selva Prabhakaran. (2019, February 18). *Time Series*. Retrieved from Machine Learning Plus: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

*Statistical capabilities*. (2018). Retrieved from Statgraphics 18: http://www.statgraphics.com/time-series-analysis-and-forecasting

WILL KENTON. (2019, July 16). *Tools For fundamental Analysis*. Retrieved from Investopedia: https://www.investopedia.com/terms/t/timeseries.asp

# Final Year Documentation

## Chapter 1: Introduction to the problem:

### Introduction:

Quantitative and qualitative techniques for forecasting help ceos and managers to develop business goals and business strategies. Forecasting can be based on historical data pattern which can be used to forecast future market behavior. In business, sound predictions of demands of goods and trends are no longer important fields. In the present time, if managers are to deal with seasonality, sudden changes in demand levels and huge swings to economy, it would become a difficult and hectic task. So Forecasting can help them deal with these difficulties. Our project include the implementation of Prophet Model which fit yearly, weekly and daily seasonality and with the holidays effect.

### Purpose:

The purpose of Time Series Sales Forecasting project is to predict future sales of stores. Sales prediction in a business give a firm idea what to expect in the coming days, months and year. By sales

forecasting a company can manage workforce and their resources. Thus the project holds a great importance in business environment

## Objectives:

- Implementing an appropriate ML model to get the output of maximum accuracy.
- Validating forecasts, in this step we will compare predicted sales to real sales of time series by using Root Mean Square Error (RMSE), RMSE tells us how far our model was able to predict sales accurately
- Producing and Visualizing forecast and displaying patterns on different time scales such as yearly, weekly and daily.

## Existing Solution:

The existing solution for the goal of sales forecasting, includes models like ARIMA, SARIMA etc but the problem with these models is that they can't handle seasonality component of time series data.

## Proposed Solution:

Our project will include a more advanced approach for forecasting i.e Prophet model which fit yearly, weekly and daily seasonality and with the holidays effect. Prophet model by facebook can effectively manage seasonality in a time series data. Thus this approach is better than the old approaches.

## References:

(n.d.). Retrieved from https://ieeexplore.ieee.org/document/278253/references#references

(n.d.). Retrieved from https://ieeexplore.ieee.org/document/278253/references#references

(n.d.). Retrieved from https://towardsdatascience.com/the-best-forecast-techniques-or-how-to-predict-from-time-series-data-967221811981

(n.d.). Retrieved from https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique

(n.d.). Retrieved from https://medium.com/datadriveninvestor/systematic-solution-for-time-series-forecasting-in-real-business-problems-2816747799d6

# Chapter 2: Software Requirement Specification:

## Introduction:

### Purpose:

Sales are the backbone of business. Accurate Sales predictions allow companies to make informed business decisions and allow them to analyse long term and short term performance measure. Old solutions are not effective in making predictions, but advancements are made to solve these problems in a more effective way. Business companies need these modern technologies to better solve their problems. Thus this project aims to use modern approaches in order to make predictions in a more effective way. Our product will be guessing the future sales on a daily, weekly and yearly basis.

## System Overview:

Infrastructure of this project includes an application where user can perform different operations, also user can view the predicted sales.Application has the functionality to import data from a CSV file or Excel file. Then it will process this date to make predictions and finally the predicted sales will be shown on the application. Other operation includes editing of user profile, performing search operation in data, sorting the data by date.

## Definition, acronyms and abbreviations:

| **Term** | **Explanation** |
|---|---|
| CSV | Comma Separated Values |
| GUI | Graphical User Interface |
| SRS | Software Requirement Specification |
| ERD | Entity Relationship Diagram |

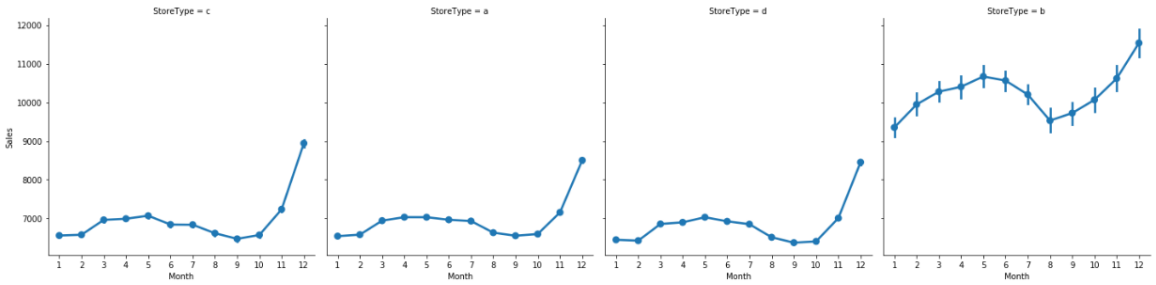
## Assumption and Dependencies:

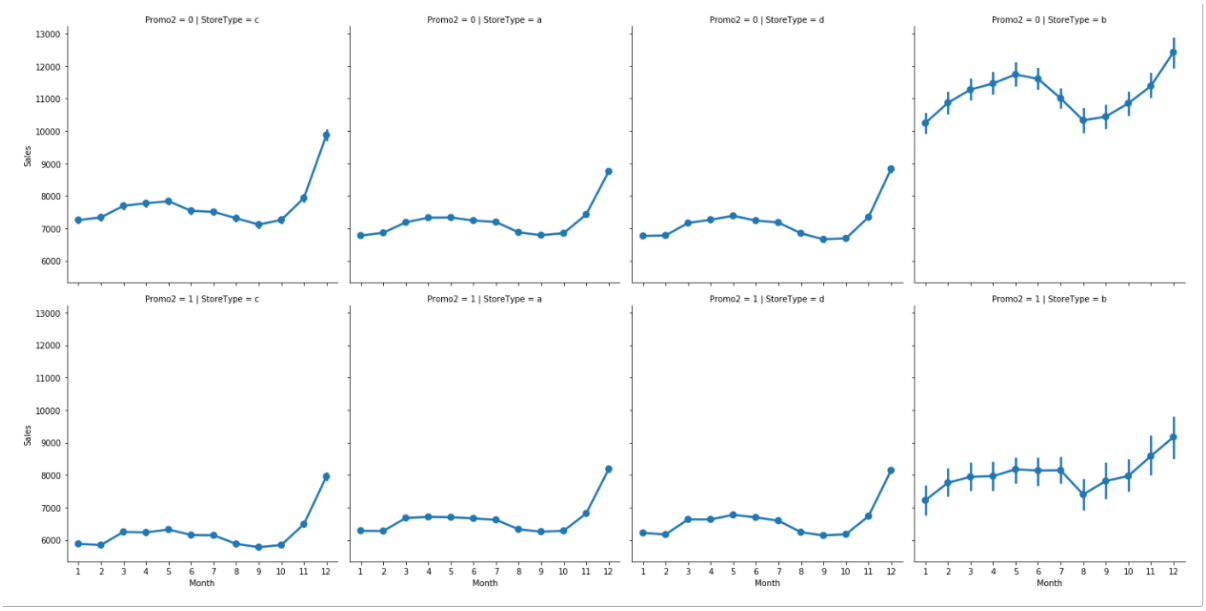
- We assume that time series data have seasonal components
- We assume that the internet connection is established in order to use our product
- We assume that data to import to application should be in specified format so that error may not occur in processing

# Interfaces:

## User Interface:

User interface of application will be easy to use and attractive. Basically it will consist of four pages Login, Administration, Dashboard, Report and Home page. Features and Characteristics of application user interface will be as follows, interface or results are shown on web browsers. Administration page will provide the functionality of adding new files which are CSV files containing date and sales fields. Login page to enter credentials in order to access database. Report page provide the results i.e predicted sales in graphical format or tabular format on a daily or yearly basis.






Thus giving a more attractive visualization and user will have a quick intuition towards rise or fall of sales.

### Hardware Interfaces:

- Our hardware interfaces include development computer having the capability to tune the models
- May require Virtual Private server for processing of the data

### Software Interfaces:

- Web Browser such as Google Chrome or Firefox
- Database Server suchs as MySQL
- Web Server such as Flask 2.8 or Django 2.0
- Development end includes HTML5, Bootstrap
- Python, Scikit, Numpy and Matplotlib of latest and stable versions

## Specific requirements:

### Functional Requirements:

- Forecasting:

  Graphical and Tabular Representation of the future sales predicted by the model on a yearly.

- Seasonality:

  Model will be able to handle seasonality or other recurring events in the data to make predictions.

- Holidays effect:

  If there are holidays in the dataset it would be able to model those holidays to make prediction with holiday effect

- File import:

  Functionality to import file from computer to application for processing to make predictions

- Reports:

  Generating reports and showing up predictions in one page.

### Non Functional Requirements:

- Usability:

Since application design will focus on simplicity and provide information about each operation to enhance usability. Graphs on screen will be simplistic and clear.

- Performance:

Forecasting sales process may or may not consume time depend upon the system specification. Performance requirements include that product should be able receive and generate forecasting in just a few seconds.

- Supportability:
  a. Code will be written in according with PEP8 standards.
  b. UpperCamelCase for classes' names
  c. A single line will contain no more than 79 characters
  d. File encoding will be in UTF-8
  e. Several built in or other libraries will be used for analyzing and visualizing the data

## Use Case Analysis:

### UseCase1: Login

1. Alternative Course:

   If the user get an error while log in process, the user must check if he has registered His/her id or not. In addition to this, the user should make sure if he is entering the correct password or not. If the id registered and the password is also correct then there must be some internet connection problem or server may be busy at that time so the user may try again later.

2. Pre-Conditions:

   User need to place his/her email and password in order to run Application.

3. Post-Conditions:

   After logging in, user is provided to access the application and he/she can predict sales and expenses using application. Also there is a signup option where a new user can enter. His/her self. Login has also another option "password verification", where we will verified our password. Sign up data will be saved in database.

4. Success Criteria:

   Now the user can easily log in the application and use the application whenever he wants. Users log in once and get one click access to all the resources they need to do their jobs. New user can add his data using sign up option where they will be username, email, password and phone number to make his ID for future use.

5. Primary Actor:
    i. Customer

6. Secondary Actor:
    i. Admin
    ii. Database

UseCase2: Producing and visualization

1. Alternative Course:

When user will put data as csv file in the Producing and visualization then if there is any error in uploading the data file, the user must check the file format of data. In addition to this, user must check the file format either data is in csv, excel format or not.

2. Pre-Conditions:

User need to add his/her data file.

3. Post-Conditions:

After putting data (file) in the system, the application will predict the daily, monthly and yearly sales. There is also another secondary actor database in which all the data will be saved.

4. Success Criteria:

Forecasting model are the mathematical models which are makes certain assumptions about the time series, captures certain features and the predicts the next unknown values, for Example Decomposition of Time Series is one the most commonly used model in Time series forecast Model. Our model clearly captured sales of future seasonality. As we forecast further out into the future out into the future, it is natural for us to become more less confident in our values. This is reflected by the intervals generated by our model, which grow larger as we move further into the future.

5. Primary Actor:
    i. Customer

6. Secondary Actor:
    i. Admin
    ii. Database

UseCase3: Trend and Forecast Visualization

a. Alternative Course:

In order to deal with Trend and forecast visualization we will have some sort of past data from which we will predict the future expenses. For example we will have the past data of product and we will also check for trending items, we will analyze the data and predict the expenses which the user will face in future. If there will be some sort of application error etc.

then we will check the backend used functionalities or we will restart the app to get rid of this problem.

b. Pre-Conditions:

Firstly User need to predict the expense of product, he will analyze the data.

c. Post-Conditions:

After doing this action there will be an option for trending items also. All the data will be saved in database also.

d. Success Criteria:

The Seasonal component models patterns of change in a time series within a year. These patterns relate to periodic fluctuations of constant length, tend to repeat themselves each year. The Trend represents changes in the level of the series. The Cyclic component refers to patterns, or waves, in the data that are repeated after approximately equal intervals with approximately equal intensity, with period normally larger than seasonal period. Usually the trend and cyclic component are together treated as the Trend component.

e. Primary Actor:
   i. Customer

f. Secondary Actor:
   i. Admin
   ii. Database

### UseCase4: Graphical User Interface

1. Alternative Course:

We are giving web Application interface to this system (Time Series Forecasting), if user will face some sort of difficulties in interface, then we have other options to make different and well form graphical interface like android application.

2. Pre-Conditions:

We have use case for GUI, customer is interacting with use case directly, GUI will have some forms and tables of data.

3. Post-Conditions:

After Putting the data into forms, user will get response from system.

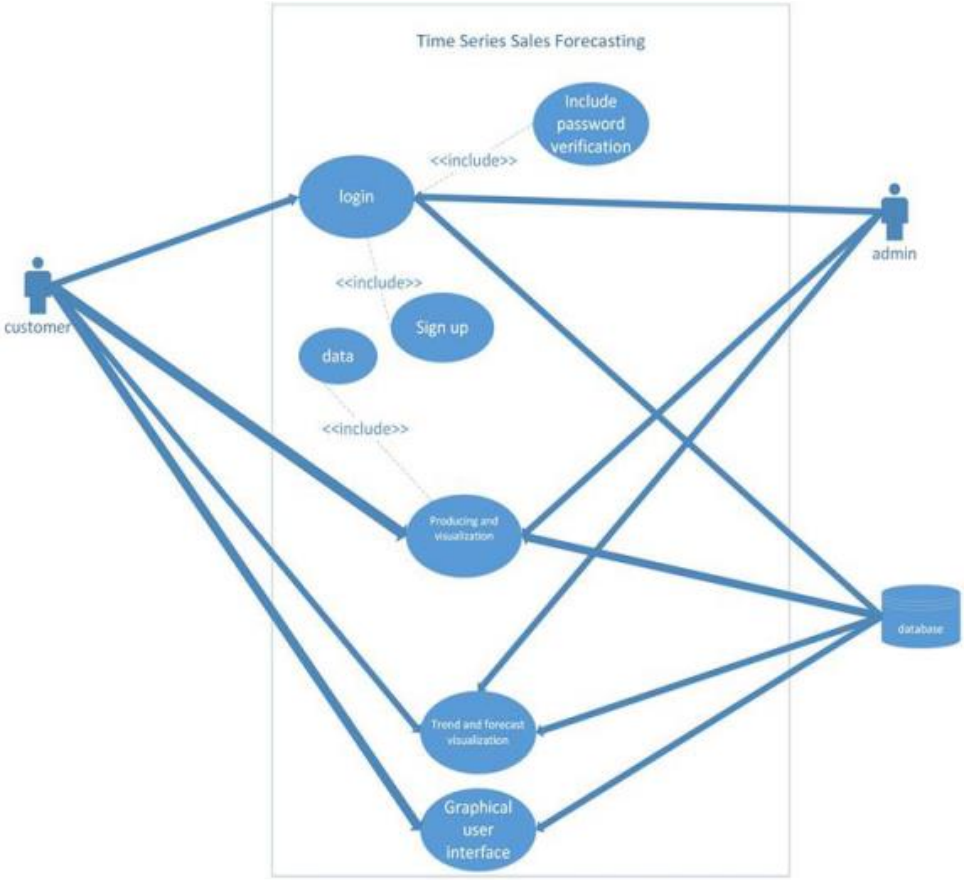
4. Success Criteria:

Now user can see results of his data, means the data user added to the system to get future sales prediction, he will get that visual result on screen. This GUI generally will provide users with immediate, visual feedback about the effect of each action and allows

multiple programs and instances to be displayed simultaneously and user do not need to know any programming languages.

5. Primary Actor:
    a. Customer

6. Secondary Actor:
    a. Admin
    b. Database




## Entity Relationship Diagram:

LOGIN

| Field | Data Type | Description | Constraint |
|---|---|---|---|
| LoginID | INT(11) unique | Unique number to identify the User | AUTO_INCREMENT Primary Key |
| `SignUpID | INT(11) | PK from SIGNUP table | Foreign Key |
| Email | VARCHAR(30) | Email address of User | |

| | | | |
|---|---|---|---|
| Password | VARCHAR(20) | User password need for login | |

SIGNUP

| Field | Data Type | Description | Constraint |
|---|---|---|---|
| SignUpID | INT(11) Unique | number to identify the User | AUTO_INCREMENT Primary Key |
| FirstName | VARCHAR(30) | First name of User | |
| LastName | VARCHAR(30) | Last name of User | |
| CNIC | INT(15) | CNIC number of user | |
| Address | VARCHAR(30) | Home Address of user | |
| PhoneNo | INT(15) | Phone Number of user | |

Store

| Field | Data Type | Description | Constraint |
|---|---|---|---|
| Store_ID | INT(11) Unique | Number to identify the store | AUTO_INCREMENT Primary Key |
| StoreName | VARCHAR(30) | Name of Store | |
| Address | VARCHAR(30) | Address of Store | |
| ContactNo | VARCHAR(30) | Contact number of Store | |
| Login_ID | INT(11) | PK from LOGIN table | Foreign Key |

Items

| Field | Data Type | Description | Constraint |
|---|---|---|---|
| Item_ID | INT(11) unique | Number to identify items | AUTO_INCREMENT Primary Key |
| itemName | VARCHAR(20) | Name of an item | |

Sales

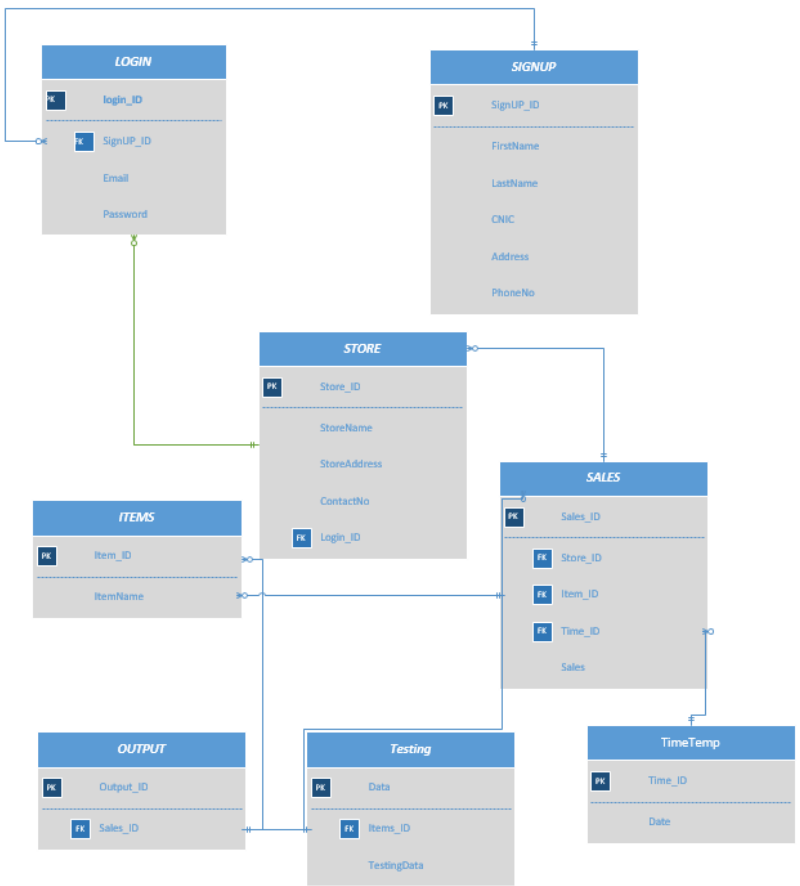
| Field | Data Type | Description | Constraint |
|---|---|---|---|
| Sales_ID | INT(11) unique | Number to identify sales | AUTO_INCREMENT Primary Key |
| Store_ID | INT (11) | PK from Store table | Foreign Key |
| Items_ID | INT (11) | PK from Items table | Foreign Key |
| Time_ID | INT (11) | PK from Time table | Foreign Key |
| Sales | VARCHAR(20) | Different item sales prediction data | |

Testing

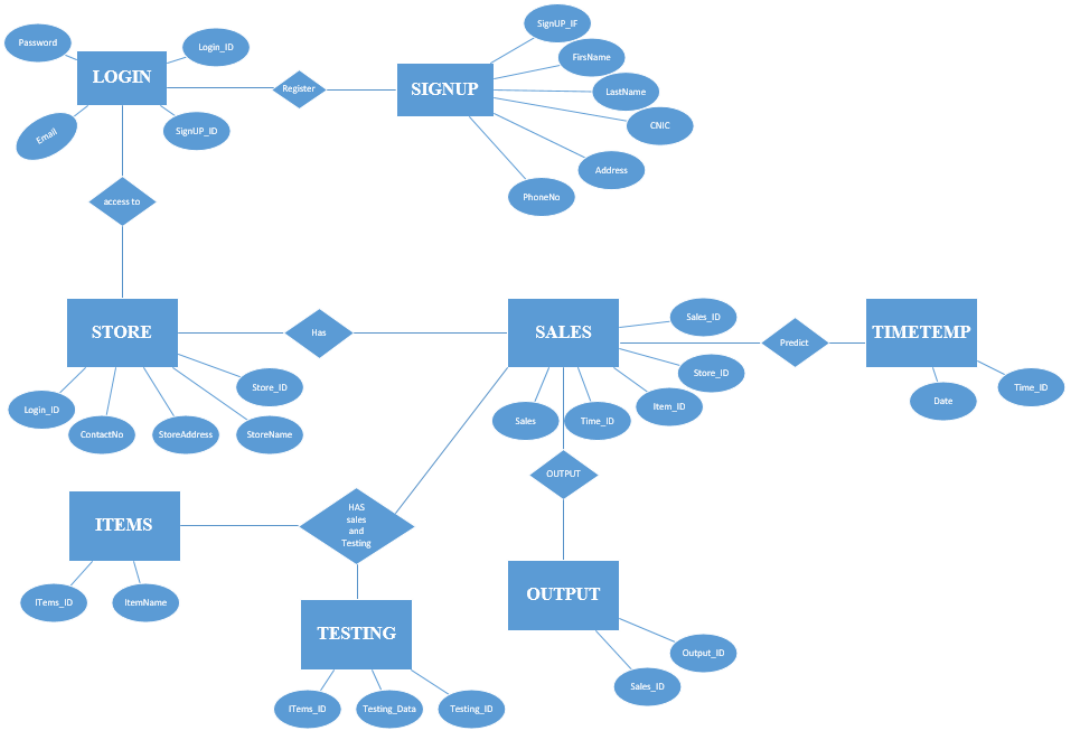
| Field | Data Type | Description | Constraint |
|---|---|---|---|
| Testing_ID | INT(11) unique | Number to identify Testing | AUTO_INCREMENT Primary Key |
| Items_ID | INT (11) | PK from Items table | Foreign Key |
| TestingData | VARCHAR | Data which is to be test | |

Output

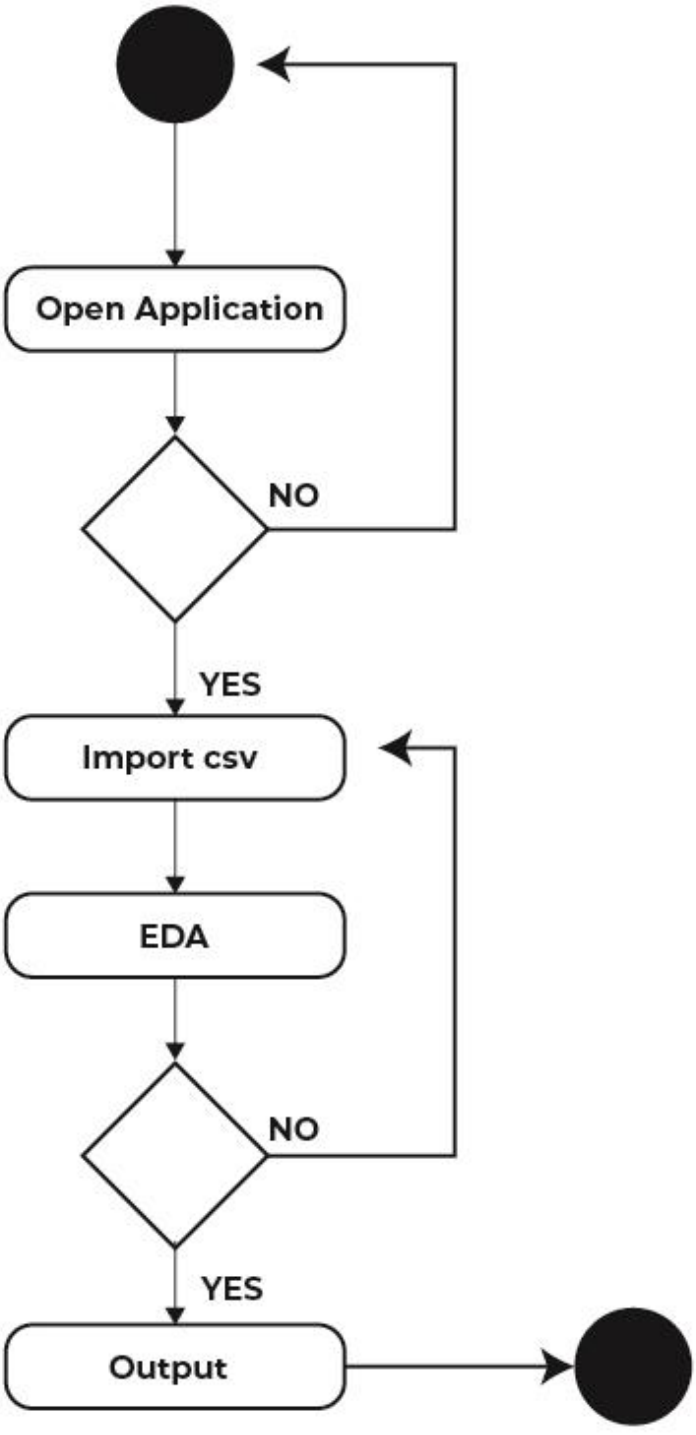
| Field | Data Type | Description | Constraint |
|---|---|---|---|
| Output_ID | INT(11) unique | Number to identify Output | AUTO_INCREMENT Primary Key |
| Sales_ID | INT (11) | PK from Sales table | Foreign Key |



## Data Flow Diagram:




## Activity Diagram:




Description:

As we can see in above figure firstly we will open the application, using login, if login is not successful then it will goes again to open application activity. If login is successful, then will import the comma separated file csv, then will perform EDA for clear the data, having no space and extra commas. If EDA performed will we will get our output.

- **Activities:**

| Open Application |
|---|

| |
|---|
| Import CSV |
| EDA |
| Output |