Project Brief

This project is inspired by the original Kaggle competition found on this website: https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data. The datasets used in this project are also curated from the original Kaggle data. However, this project is not an attempt to solve the problems provided by the competition.

This project basically describes how to design a database and the integration of a sqlite database with python with the use of a GUI such as Tkinter.

Walmart operates a large number of retail stores in the USA. They want to predict sales in different departments from various factors: such as the temperature, whether there is a holiday in specific week, the unemployment rate, markdown on the prices.

Walmart have run a successful trial of their machine learning for some of their stores in California. Because only a trial was run the data was collected using csv files:

- Features_data_set.csv
- sales_data-set.csv
- stores_data-set.csv
- store_info.csv

The data set contains information organized by week. The file **Features_data_set.csv** contains information such as: Date, temperature, CPI (Consumer Price Index), Unemployment (% of the workforce who are unemployed.)
Information about the sales from each department in each store is in the **sales_dataset.csv** file. The **store_info.csv** file contains information about the store, such as the name of the manager and the address. The **stores_data-set.csv** file also contains information about the stores.

In this project, a database is designed to store the data in the csv files. The database design includes an entity relation diagram, using the crowsfoot notation. The database is designed to follow the third normal form (3NF). A SQL schema for all the tables is also produced.

A python script is written to read in the csv files and populate a relational database such as SQLlite. Further, a GUI interface is written using Tkinter to update the manager of a specific store to the database. The code checks that the email is in the correct format: xxx.yyy@Walmart.org where xxx and yyy are strings of letters. An additional GUI to calculate the mean store size based on type as well as produce a time-series sales graph of any specific store and department is also provided.

Note: If you want to run the codes in the .ipynb file provided in this repository, please ensure downloading all the csv files in this repository and placing them in the same directory as the **walmart_db.ipynb** file.