AIML Capstone
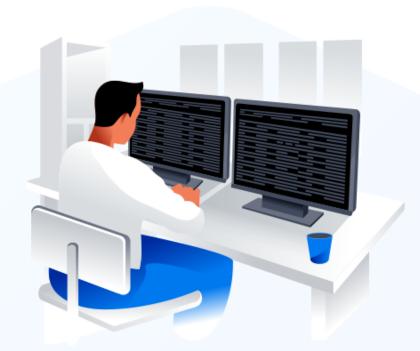
Sales Forecasting

Business Scenario

Problem statement:

In ever-changing competitive market conditions, there is a need to make correct decisions and plans for future events related to business like sales, production, and many more. The effectiveness of a decision taken by business managers is influenced by the accuracy of the models used. Demand is the most important aspect of a business's ability to achieve its objectives. Many decisions in business depend on demand, like production, sales, and staff requirements. Forecasting is necessary for business at both international and domestic levels.

Business Scenario

Problem objective:

Fresh Analytics, a data analytics company, aims to comprehend and predict the demand for various items across restaurants. The primary goal of the project is to determine the sales of items across different restaurants over the years.

Dataset Snapshot

Name of the dataset: **restaurants.csv**

id	name
1	Bob's Diner
2	Beachfront Bar
3	Sweet Shack
4	Fou Cher
5	Corner Cafe
6	Surfs Up

Dataset Snapshot

Name of the dataset: **sales.csv**

date	item_id	price	item_count
########	3	29.22	2
########	4	26.42	22
########	12	4.87	7
########	13	4.18	12
########	16	3.21	136
########	19	2.89	108
########	21	23.23	13
########	23	10.86	10
########	27	26.21	1
########	35	4.36	1
########	42	2.91	1
########	45	23.43	4
########	56	28.75	18
########	59	23.37	30
########	67	7	23
########	73	5.16	2
########	75	5.6	1
########	76	17.93	32
########	90	7.71	2

Dataset Snapshot

Name of the dataset: **Items.csv**

id	store_id	name	kcal	cost
1	4	Chocolate	554	6.71
2	4	Breaded F	772	15.09
3	1	Sweet Fru	931	29.22
4	1	Amazing S	763	26.42
5	5	Milk Cake	583	6.07
6	4	Fruity Fro	537	5.61
7	4	Fruity Mil	605	8.1
8	6	Amazing T	258	24.98
9	1	Orange Ju	135	3.91
10	5	Mutton w	820	21.13
11	4	Fantastic	787	19.48
12	1	Fantastic	478	4.87
13	1	Sweet Fro	490	4.18
14	6	Amazing C	472	3.9
15	5	Original La	380	18.78
16	1	Frozen Mi	284	3.21
17	4	Awesome	581	6.21
18	1	Mutton Lu	566	18.82
19	1	Strawberr	145	2.89

Dataset Description

restaurants.csv: This data provides information about the restaurants or store.

Variables	Description
id	Unique identification of the restaurant or store
name	Name of the restaurant

Dataset Description

items.csv: This data contains information about items.

Variables	Description
id	Unique identification of the item
store_id	Unique identification of the store
name	Name of the item
kcal	A measure of energy nutrients (calories) in the item
cost	The unit price of the item bought

Dataset Description

sales.csv: This contains the count of a particular item sold at a particular store or restaurant for different dates.

Variables	Description
date	Date of purchase
item	Name of the item bought
Price	Unit price of the item
item_count	Total count of the items bought on that day

Project Task

Preliminary analysis:

   a.   Import the datasets into the Python environment

   b.   Examine the dataset's shape and structure, and look out for any outlier

   c.   Merge the datasets into a single dataset that includes the date, item id, price, item count, item names, kcal values, store id, and store name

Exploratory data analysis:

   a.   Examine the overall date wise sales to understand the pattern

   b.   Find out how sales fluctuate across different days of the week

   c.   Look for any noticeable trends in the sales data for different months of the year

   d.   Examine the sales distribution across different quarters averaged over the years. Identify any noticeable patterns.

   e.   Compare the performances of the different restaurants. Find out which restaurant had the most sales and look at the sales for each restaurant across different years, months, and days.

Project Task

f. Identify the most popular items overall and the stores where they are being sold. Also, find out the most popular item at each store

g. Determine if the store with the highest sales volume is also making the most money per day

h. Identify the most expensive item at each restaurant and find out its calorie count

Forecasting using machine learning algorithms:

a. Build and compare linear regression, random forest, and XGBoost models for predictions

- Generate necessary features for the development of these models, like day of the week, quarter of the year, month, year, day of the month and so on

- Use the data from the last six months as the testing data

- Compute the root mean square error (RMSE) values for each model to compare their performances

- Use the best-performing models to make a forecast for the next year

Project Task

Forecasting using deep learning algorithms:

a. Use sales amount for predictions instead of item count

b. Build a long short-term memory (LSTM) model for predictions

- Define the train and test series

- Generate synthetic data for the last 12 months

- Build and train an LSTM model

- Use the model to make predictions for the test data

c. Calculate the mean absolute percentage error (MAPE) and comment on the model's performance

d. Develop another model using the entire series for training, and use it to forecast for the next three months

Thank You