## FINAL PROJECT SUBMISSION

## **Smaira**

23112314

# **Objectives**

- To Familiarise Project Planning and Deployment Strategies
- To have an understanding of Technical Components contributing to an analytics-based project
- To create an analysis involving KDD Process, Data Analytics & Predictive Modelling using Python and other practices as discussed during the Python Sessions
- To integrate good coding practices while deducing solutions to challenges

```
In [1]: import pandas as pd
        from sklearn.model_selection import train_test_split
        from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error
        import seaborn as sns
        from sklearn.preprocessing import LabelEncoder
        from sklearn.preprocessing import MinMaxScaler
        from sklearn.linear_model import LogisticRegression
        from sklearn.metrics import accuracy_score
        import warnings as w
        import numpy as np
        import matplotlib.pyplot as plt
        from sklearn.metrics import roc curve, auc
        from sklearn.preprocessing import label binarize
        from sklearn.metrics import confusion_matrix
        w.filterwarnings('ignore')
In [2]: df=pd.read csv(r"C:\Users\smair\Downloads\CIA 3 Data - Dinner - Smaira.csv")
```

In [3]: **df** 

$\cap$	[3]	
out	Lフ」	۰

		Register Number	Name	Date	Food Type	Location	Where in College?	Variety	Avg money	Dinner time
	0	23112301	Arghya Basu	4.3	Veg	Market	Not in college	Italian	₹100- ₹150	7pm- 8pm
	1	23112302	Aryan Singhal	4.3	Non - Veg	College	Steaming Mugs	Chinese	₹200- ₹250	6pm- 7pm
	2	23112303	Avinash Ruthvik	4.3	Non - Veg	Market	Not in college	Indian	₹100- ₹150	9pm- 10pm
	3	23112305	Harshil Mishra	4.3	Non - Veg	Market	Not in college	Chinese	₹150- ₹200	6pm- 7pm
	4	23112306	Juhi Rathore	4.3	Veg	Home- Cooked	Not in college	Chinese	₹50- ₹100	7pm- 8pm
	•••									
7	94	23112316	Tarit Kumar Singh	1.6	Veg	College	Cafe by the Valley	Chinese	₹100- ₹150	9pm- 10pm
7	95	23112317	Urvi Saran	1.6	Non - Veg	Market	Not in college	Indian	₹200- ₹250	9pm- 10pm
7	96	23112318	Veda Yogesh Rane	1.6	Veg	College	Cafe by the Valley	Indian	₹150- ₹200	Past 10pm
7	97	23112320	Abhimanyu Sharma	1.6	Veg	College	Chopsticks	Chinese	₹50- ₹100	7pm- 8pm
7	98	23112321	Kratik Lodha	1.6	Veg	Home- Cooked	Not in college	Indian	₹50- ₹100	9pm- 10pm

799 rows × 9 columns

```
df['Food Type']
In [4]:
                      Veg
Out[4]:
        1
                Non - Veg
         2
                Non - Veg
         3
                Non - Veg
         4
                      Veg
        794
                      Veg
        795
                Non - Veg
        796
                      Veg
        797
                      Veg
        798
                      Veg
        Name: Food Type, Length: 799, dtype: object
In [5]:
         df
```

011+[5]

	Register Number	Name	Date	Food Type	Location	Where in College ?	Variety	Avg money	Dinner time
0	23112301	Arghya Basu	4.3	Veg	Market	Not in college	Italian	₹100- ₹150	7pm- 8pm
1	23112302	Aryan Singhal	4.3	Non - Veg	College	Steaming Mugs	Chinese	₹200- ₹250	6pm- 7pm
2	23112303	Avinash Ruthvik	4.3	Non - Veg	Market	Not in college	Indian	₹100- ₹150	9pm- 10pm
3	23112305	Harshil Mishra	4.3	Non - Veg	Market	Not in college	Chinese	₹150- ₹200	6pm- 7pm
4	23112306	Juhi Rathore	4.3	Veg	Home- Cooked	Not in college	Chinese	₹50- ₹100	7pm- 8pm
•••									
794	23112316	Tarit Kumar Singh	1.6	Veg	College	Cafe by the Valley	Chinese	₹100- ₹150	9pm- 10pm
795	23112317	Urvi Saran	1.6	Non - Veg	Market	Not in college	Indian	₹200- ₹250	9pm- 10pm
796	23112318	Veda Yogesh Rane	1.6	Veg	College	Cafe by the Valley	Indian	₹150- ₹200	Past 10pm
797	23112320	Abhimanyu Sharma	1.6	Veg	College	Chopsticks	Chinese	₹50- ₹100	7pm- 8pm
798	23112321	Kratik Lodha	1.6	Veg	Home- Cooked	Not in college	Indian	₹50- ₹100	9pm- 10pm

799 rows × 9 columns

#### In [6]: df.describe()

Out[6]:		Register Number	Date
	count	7.990000e+02	799.000000
	mean	2.311231e+07	14.930038
	std	6.314041e+00	7.915728
	min	2.311230e+07	1.400000
	25%	2.311231e+07	8.300000
	50%	2.311231e+07	15.300000
	75%	2.311232e+07	21.300000
	max	2.311232e+07	30.500000

Out[8]:

	Register Number	Name	Date	Food Type	Location	Where in College?	Variety	Avg money	Dinner time
0	23112301	Arghya Basu	asu 4.3 Veg		Market	Not in college	Italian	₹100- ₹150	7pm- 8pm
1	23112302	Aryan Singhal	4.3	Non - Veg	College	Steaming Mugs	Chinese	₹200- ₹250	6pm- 7pm
2	23112303	Avinash Ruthvik	4.3	Non - Veg	Market	Not in college	Indian	₹100- ₹150	9pm- 10pm
3	23112305	Harshil Mishra	4.3	Non - Veg	Market	Not in college	Chinese	₹150- ₹200	6pm- 7pm
4	23112306	Juhi Rathore	4.3	Veg	Home- Cooked	Not in college	Chinese	₹50- ₹100	7pm- 8pm
•••									
794	23112316	Tarit Kumar Singh	1.6	Veg	College	Cafe by the Valley	Chinese	₹100- ₹150	9pm- 10pm
795	23112317	Urvi Saran	1.6	Non - Veg	Market	Not in college	Indian	₹200- ₹250	9pm- 10pm
796	23112318	Veda Yogesh Rane	1.6	Veg	College	Cafe by the Valley	Indian	₹150- ₹200	Past 10pm
797	23112320	Abhimanyu Sharma	1.6	Veg	College	Chopsticks	Chinese	₹50- ₹100	7pm- 8pm
798	23112321	Kratik Lodha	1.6	Veg	Home- Cooked	Not in college	Indian	₹50- ₹100	9pm- 10pm

799 rows × 9 columns

```
In [9]:
        df.keys()
         Index(['Register Number ', 'Name', 'Date ', 'Food Type', 'Location',
Out[9]:
                'Where in College ?', 'Variety', 'Avg money', 'Dinner time'],
               dtype='object')
In [10]: df
```

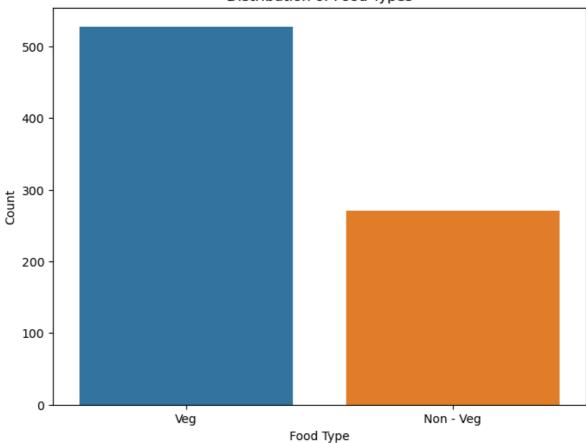
Out[10]:

						,						
		Register Number	Name Date		Location	Where in College?	Variety	Avg money	Dinner time			
	0	23112301	Arghya Basu	4.3	Veg	Market	Not in college	Italian	₹100- ₹150	7pm- 8pm		
	1	23112302	Singhal		Non - Veg	College	Steaming Mugs	Chinese	₹200- ₹250	6pm- 7pm		
	2	23112303	Avinash Ruthvik	4.3	Non - Veg	Market	Not in college	Indian	₹100- ₹150	9pm- 10pm		
	3	23112305	Harshil Mishra	4.3	Non - Veg	Market	Not in college	Chinese	₹150- ₹200	6pm- 7pm		
	4	23112306	Juhi Rathore	4.3	Veg	Home- Cooked	Not in college	Chinese	₹50- ₹100	7pm- 8pm		
	•••											
79	94	23112316	Tarit Kumar Singh	1.6	Veg	College	Cafe by the Valley	Chinese	₹100- ₹150	9pm- 10pm		
79	95	23112317	Urvi Saran	1.6	Non - Veg	Market	Not in college	Indian	₹200- ₹250	9pm- 10pm		
79	96	23112318	Veda Yogesh Rane	1.6	Veg	College	Cafe by the Valley	Indian	₹150- ₹200	Past 10pm		
79	97	23112320	Abhimanyu Sharma	1.6	Veg	College	Chopsticks	Chinese	₹50- ₹100	7pm- 8pm		
79	98	23112321	Kratik Lodha	1.6	Veg	Home- Cooked	Not in college	Indian	₹50- ₹100	9pm- 10pm		

#### 799 rows × 9 columns

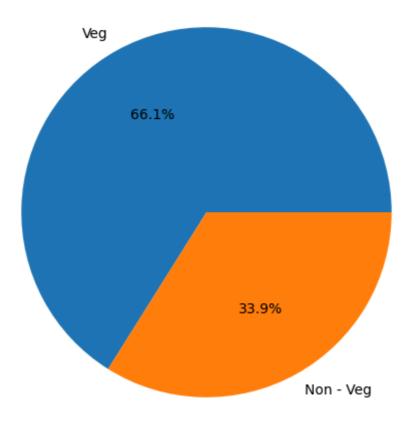
```
In [11]: plt.figure(figsize=(8, 6))
    sns.countplot(data=df, x='Food Type')
    plt.title('Distribution of Food Types')
    plt.xlabel('Food Type')
    plt.ylabel('Count')
    plt.show()
```

#### Distribution of Food Types



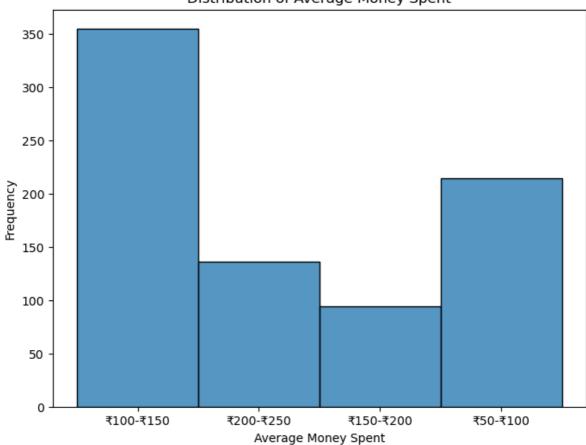
```
In [12]: plt.figure(figsize=(8, 6))
    df['Food Type'].value_counts().plot(kind='pie', autopct='%1.1f%%')
    plt.title('Distribution of Food Types')
    plt.ylabel('')
    plt.show()
```

### Distribution of Food Types

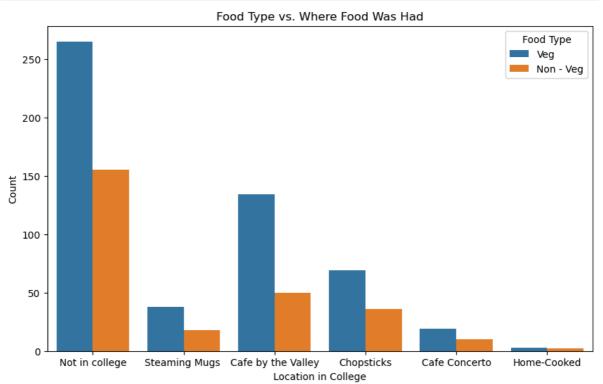


```
In [13]: plt.figure(figsize=(8, 6))
    sns.histplot(data=df, x='Avg money', bins=10)
    plt.title('Distribution of Average Money Spent')
    plt.xlabel('Average Money Spent')
    plt.ylabel('Frequency')
    plt.show()
```

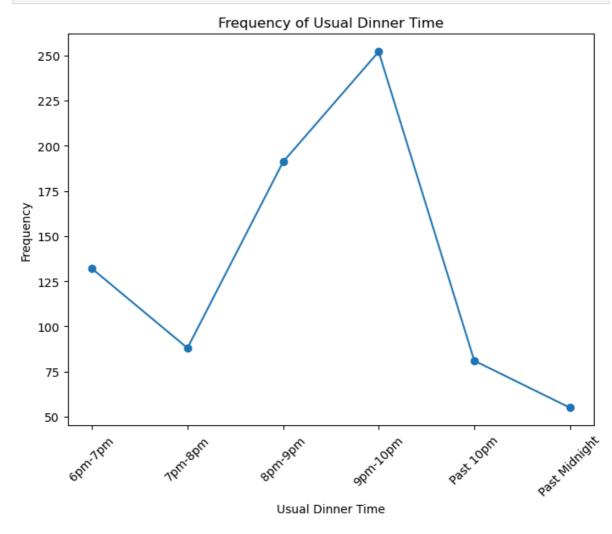
#### Distribution of Average Money Spent



```
In [14]:
    plt.figure(figsize=(10, 6))
    sns.countplot(data=df, x='Where in College ?', hue='Food Type')
    plt.title('Food Type vs. Where Food Was Had')
    plt.xlabel('Location in College')
    plt.ylabel('Count')
    plt.legend(title='Food Type')
    plt.show()
```



```
In [15]:
    plt.figure(figsize=(8, 6))
    df['Dinner time'].value_counts().sort_index().plot(kind='line', marker='o')
    plt.title('Frequency of Usual Dinner Time')
    plt.xlabel('Usual Dinner Time')
    plt.ylabel('Frequency')
    plt.xticks(rotation=45)
    plt.show()
```



```
In [20]: df.columns = df.columns.str.strip()
    df
```

Out[20]: Register Dinner **Food** Where in Avg Name Date Location Variety Number Туре College? time money ₹100-7pm-Not in 0 23112301 Arghya Basu Market Italian 4.3 Veg college ₹150 8pm ₹200-Aryan Non -Steaming 6pm-1 23112302 4.3 College Chinese Singhal ₹250 Veg Mugs 7pm ₹100-9pm-Avinash Not in Non -4.3 2 23112303 Market Indian Ruthvik college ₹150 10pm Veg Harshil Non -₹150-6pm-Not in 4.3 Chinese 3 23112305 Market Mishra college ₹200 7pm Veg Home-₹50-7pm-Not in 4 23112306 Juhi Rathore 4.3 Veg Chinese Cooked college ₹100 8pm Tarit Kumar ₹100-9pm-Cafe by the 23112316 794 1.6 Veg College Chinese Valley ₹150 10pm Singh Not in ₹200-9pm-Non -795 23112317 Urvi Saran 1.6 Market Indian ₹250 10pm college Veg ₹150-Past Veda Yogesh Cafe by the 23112318 796 1.6 College Indian Veg Rane Valley ₹200 10pm ₹50-7pm-Abhimanyu 797 23112320 Chopsticks 1.6 Veg College Chinese Sharma ₹100 8pm Home-Not in ₹50-9pm-798 23112321 Kratik Lodha Indian 1.6 Veg ₹100 Cooked college 10pm

799 rows × 9 columns

```
In [21]: if 'Date' in df.columns:
    df['Date'] = df['Date'].astype("str")
    df['Date'] = pd.to_datetime(df['Date'], format='%d.%m')

    df['Date'] = df['Date'].dt.strftime('%Y-%m-%d').str.replace('1900', '2024')

    df['Weekday'] = pd.to_datetime(df['Date']).dt.strftime('%A')

    df
else:
    print("Column 'Date' not found in the DataFrame.")
In [22]: df
```

Out[22]:

	Register Number	Name	Date	Food Type	Location	Where in College ?	Variety	Avg money	Dinner time	Weekday
0	23112301	Arghya Basu	2024- 03-04	Veg	Market	Not in college	Italian	₹100- ₹150	7pm- 8pm	Monday
1	23112302	Aryan Singhal	2024- 03-04	Non - Veg	College	Steaming Mugs	Chinese	₹200- ₹250	6pm- 7pm	Monday
2	23112303	Avinash Ruthvik	2024- 03-04	Non - Veg	Market	Not in college	Indian	₹100- ₹150	9pm- 10pm	Monday
3	23112305	Harshil Mishra	2024- 03-04	Non - Veg	Market	Not in college	Chinese	₹150- ₹200	6pm- 7pm	Monday
4	23112306	Juhi Rathore	2024- 03-04	Veg	Home- Cooked	Not in college	Chinese	₹50- ₹100	7pm- 8pm	Monday
•••										
794	23112316	Tarit Kumar Singh	2024- 06-01	Veg	College	Cafe by the Valley	Chinese	₹100- ₹150	9pm- 10pm	Saturday
795	23112317	Urvi Saran	2024- 06-01	Non - Veg	Market	Not in college	Indian	₹200- ₹250	9pm- 10pm	Saturday
796	23112318	Veda Yogesh Rane	2024- 06-01	Veg	College	Cafe by the Valley	Indian	₹150- ₹200	Past 10pm	Saturday
797	23112320	Abhimanyu Sharma	2024- 06-01	Veg	College	Chopsticks	Chinese	₹50- ₹100	7pm- 8pm	Saturday
798	23112321	Kratik Lodha	2024- 06-01	Veg	Home- Cooked	Not in college	Indian	₹50- ₹100	9pm- 10pm	Saturday

799 rows × 10 columns

df.head() In [23]: Out[23]: Register Where in Dinner Food Avg Name Date Location Variety Weekday College? Number Туре money time Arghya 2024-₹100-Not in 7pm-23112301 0 Market Italian Monday Veg Basu 03-04 college ₹150 8pm Aryan 2024-Non Steaming ₹200-6pm-23112302 College Chinese Monday 1 Singhal 03-04 - Veg Mugs ₹250 7pm 2024-₹100-Avinash Non Not in 9pm-23112303 Market Indian Monday Ruthvik 03-04 - Veg college ₹150 10pm Harshil 2024-Non Not in ₹150-6pm-Market 3 23112305 Chinese Monday Mishra 03-04 college ₹200 - Veg 7pm 2024-Home-₹50-Juhi Not in 7pm-23112306 Veg Chinese Monday Rathore 03-04 Cooked college ₹100 8pm

In [24]: df.tail()

data anlytics cia3 27/03/2025, 22:49

Out[24]: Register Where in Dinner **Food** Avg Name Date Location Variety Weekday Number Туре College? time money Tarit Kumar 2024-Cafe by ₹100-9pm-**794** 23112316 Chinese Veg College Saturday 06-01 Singh the Valley ₹150 10pm 2024-₹200-Non Not in 9pm-795 23112317 Urvi Saran Market Indian Saturday - Veg 06-01 college ₹250 10pm Veda 2024-Cafe by ₹150-Past **796** 23112318 Indian Saturday Yogesh Veg College 06-01 the Valley ₹200 10pm Rane 2024-₹50-Abhimanyu 7pm-**797** 23112320 Chopsticks Saturday Veg College Sharma 06-01 ₹100 8pm Kratik 2024-Home-Not in ₹50-9pm-**798** 23112321 Indian Veg Saturday Lodha 06-01 ₹100 10pm Cooked college df['Date'] = pd.to\_datetime(df['Date']) In [25]: date\_cutoff = pd.to\_datetime('2024-04-15') df = df[df['Date'] <= date\_cutoff]</pre>

In [26]: df

Out[26]:

	Register Number	Name	Date	Food Type	Location	Where in College ?	Variety	Avg money	Dinner time	Weekday
0	23112301	Arghya Basu	2024- 03-04	Veg	Market	Not in college	Italian	₹100- ₹150	7pm- 8pm	Monday
1	23112302	Aryan Singhal	2024- 03-04	Non - Veg	College	Steaming Mugs	Chinese	₹200- ₹250	6pm- 7pm	Monday
2	23112303	Avinash Ruthvik	2024- 03-04	Non - Veg	Market	Not in college	Indian	₹100- ₹150	9pm- 10pm	Monday
3	23112305	Harshil Mishra	2024- 03-04	Non - Veg	Market	Not in college	Chinese	₹150- ₹200	6pm- 7pm	Monday
4	23112306	Juhi Rathore	2024- 03-04	Veg	Home- Cooked	Not in college	Chinese	₹50- ₹100	7pm- 8pm	Monday
•••										
538	23112316	Tarit Kumar Singh	2024- 04-15	Veg	College	Cafe by the Valley	Indian	₹100- ₹150	6pm- 7pm	Monday
539	23112317	Urvi Saran	2024- 04-15	Veg	Market	Not in college	Chinese	₹150- ₹200	9pm- 10pm	Monday
540	23112318	Veda Yogesh Rane	2024- 04-15	Veg	Market	Not in college	Indian	₹50- ₹100	7pm- 8pm	Monday
541	23112320	Abhimanyu Sharma	2024- 04-15	Non - Veg	Market	Not in college	Indian	₹100- ₹150	Past 10pm	Monday
542	23112321	Kratik Lodha	2024- 04-15	Non - Veg	Market	Not in college	Indian	₹200- ₹250	6pm- 7pm	Monday

543 rows × 10 columns

Out[28]:

Where Register Food in Avg Dinner Date Variety Name Location Weekday Number Type College time money ? Arghya 2024-₹100-Not in 7pm-0 23112301 Veg Market Italian Monday Basu 03-04 college ₹150 8pm Aryan 2024-₹200-Non 6pm-Steaming 23112302 College Chinese Monday Singhal 03-04 ₹250 - Veg Mugs 7pm 2024-₹100-Avinash Non Not in 9pm-2 23112303 Market Indian Monday Ruthvik 03-04 ₹150 - Veg college 10pm Harshil 2024-₹150-6pm-Non Not in 23112305 Market Chinese Monday Mishra 03-04 ₹200 college 7pm - Veg ₹50-Juhi 2024-Home-7pm-Not in 23112306 Veg Chinese Monday Rathore 03-04 Cooked college ₹100 8pm Cafe by ₹100-Tarit Kumar 2024-6pm-23112316 Veg College the Indian Monday Singh ₹150 7pm 04-15 Valley 2024-₹150-9pm-Not in 23112317 Veg 539 Urvi Saran Market Chinese Monday 04-15 college ₹200 10pm Veda 2024-Not in ₹50-7pm-540 23112318 Market Indian Monday Yogesh Veg 04-15 ₹100 college 8pm Rane Abhimanyu 2024-₹100-Non Not in **Past 541** 23112320 Market Indian Monday Sharma 04-15 ₹150 10pm - Veg college ₹200-Kratik 2024-Non Not in 6pm-**542** 23112321 Market Indian Monday Lodha 04-15 ₹250 - Veg college 7pm

543 rows × 10 columns

```
In [29]:
          df 1 = df 1[df 1['Avg money'] != 'Above 300']
In [30]:
          df_1['Avg money'].value_counts()
                       236
         ₹100-₹150
Out[30]:
                       146
          ₹50-₹100
                        91
         ₹200-₹250
                        70
         ₹150-₹200
         Name: Avg money, dtype: int64
In [31]:
         def convert currency to mean(currency str):
              currency str = currency str.replace('₹', '').replace(',', '')
              lower, upper = map(int, currency_str.split('-'))
              return (lower + upper) / 2
          df['Avg money'] = df['Avg money'].apply(convert_currency_to_mean)
          scaler = MinMaxScaler()
          df['Avg money (Normalized)'] = scaler.fit_transform(df['Avg money'].values.reshape(
          df
```

Out[31]:

Where Register Dinner Food in Avg Date Location Variety Weekday Name Number College time Type money ? Arghya 2024-Not in 7pm-0 23112301 Italian 125.0 Veg Market Monday Basu 03-04 college 8pm Aryan 2024-Non Steaming 6pm-23112302 College Chinese 225.0 Monday Singhal 03-04 Mugs 7pm - Veg Avinash 2024-Not in Non 9pm-2 23112303 Market Indian 125.0 Monday Ruthvik 03-04 college 10pm - Veg 2024-Harshil Non Not in 6pm-23112305 Market Chinese 175.0 Monday Mishra 03-04 college - Veg 7pm 2024-Juhi Home-Not in 7pm-23112306 Chinese 75.0 Veg Monday Rathore 03-04 Cooked college 8pm Cafe by Tarit Kumar 2024-6pm-23112316 Indian 125.0 Veg College the Monday Singh 04-15 7pm Valley 2024-Not in 9pm-23112317 Urvi Saran Market Chinese 175.0 539 Veg Monday 04-15 college 10pm Veda 2024-Not in 7pm-Veg 540 23112318 Yogesh Market Indian 75.0 Monday 04-15 college 8pm Rane Abhimanyu 2024-Not in Non Past **541** 23112320 Market Indian 125.0 Monday - Veg Sharma 04-15 college 10pm Kratik 2024-Non Not in 6pm-225.0 **542** 23112321 Indian Market Monday

college

7pm

543 rows × 11 columns

Lodha 04-15

- Veg

```
In [32]: def categorize_avg_money(amount):
    if amount >= 200:
        return 'High'
    elif amount >= 100:
        return 'Medium'
    else:
        return 'Low'
    df['Avg money category'] = df['Avg money'].apply(categorize_avg_money)
    print(df)
```

```
Register Number
                                 Name
                                            Date Food Type Location \
0
           23112301
                          Arghya Basu 2024-03-04
                                                                 Market
                                                       Veg
1
           23112302
                        Aryan Singhal 2024-03-04 Non - Veg
                                                               College
           23112303 Avinash Ruthvik 2024-03-04 Non - Veg
2
                                                                 Market
3
                      Harshil Mishra 2024-03-04 Non - Veg
           23112305
                                                                 Market
4
           23112306
                          Juhi Rathore 2024-03-04
                                                       Veg Home-Cooked
           23112316 Tarit Kumar Singh 2024-04-15
538
                                                      Veg
                                                               College
539
           23112317
                           Urvi Saran 2024-04-15
                                                       Veg
                                                                 Market
540
           23112318
                    Veda Yogesh Rane 2024-04-15
                                                                 Market
                                                       Veg
541
           23112320
                    Abhimanyu Sharma 2024-04-15 Non - Veg
                                                                 Market
542
           23112321
                          Kratik Lodha 2024-04-15 Non - Veg
                                                                 Market
    Where in College ? Variety Avg money Dinner time Weekday \
        Not in college Italian
                                  125.0
                                           7pm-8pm Monday
0
1
        Steaming Mugs
                       Chinese
                                    225.0
                                              6pm-7pm Monday
                                   125.0
2
        Not in college
                       Indian
                                             9pm-10pm Monday
                                           6pm-7pm Monday
3
        Not in college Chinese
                                    175.0
4
        Not in college Chinese
                                    75.0
                                             7pm-8pm Monday
                   . . .
                                     . . .
                                                 . . .
538 Cafe by the Valley
                       Indian
                                   125.0
                                            6pm-7pm Monday
539
        Not in college Chinese
                                   175.0
                                             9pm-10pm Monday
540
        Not in college
                                    75.0
                       Indian
                                             7pm-8pm Monday
541
        Not in college
                        Indian
                                    125.0
                                           Past 10pm Monday
542
                                    225.0
                                             6pm-7pm Monday
        Not in college
                        Indian
    Avg money (Normalized) Avg money category
0
                  0.333333
                                      Medium
1
                  1.000000
                                        High
2
                  0.333333
                                      Medium
3
                  0.666667
                                      Medium
4
                  0.000000
                                        Low
                                         . . .
538
                  0.333333
                                      Medium
539
                  0.666667
                                      Medium
540
                  0.000000
                                        Low
541
                                      Medium
                  0.333333
542
                  1.000000
                                        High
[543 rows x 12 columns]
```

```
In [33]: new=df.drop(columns=['Name'])
    new
```

Out[33]:

Where Register Food in Avg Dinner Avg money Date Variety Location Weekday College Number money time (Normalized) ? 2024-Not in 7pm-0 23112301 125.0 Veg Market Italian Monday 0.333333 03-04 college 8pm 2024-Non Steaming 6pm-23112302 College Chinese 225.0 1.000000 Monday 03-04 - Veg Mugs 7pm 2024-Non Not in 9pm-2 23112303 Market Indian 125.0 Monday 0.333333 03-04 college 10pm - Veg 2024-Non Not in 6pm-23112305 Market Chinese 175.0 Monday 0.666667 03-04 college - Veg 7pm 2024-Home-Not in 7pm-23112306 Chinese 75.0 0.000000 Veg Monday 03-04 Cooked college 8pm Cafe by 2024-6pm-23112316 Indian 125.0 538 Veg College the Monday 0.333333 04-15 7pm Valley 2024-Not in 9pm-23112317 175.0 539 Veg Market Chinese Monday 0.666667 04-15 college 10pm 2024-Not in 7pm-540 23112318 Veg Market Indian 75.0 Monday 0.000000 04-15 college 8pm 2024-Non Not in Past 23112320 125.0 541 Market Indian Monday 0.333333 04-15 - Veg college 10pm 2024-Non Not in 6pm-**542** 23112321 Market Indian 225.0 1.000000 Monday 04-15 - Veg college 7pm

543 rows × 11 columns

```
In [34]:
          dinner_time_counts = df['Dinner time'].value_counts()
          print(dinner_time_counts)
          9pm-10pm
                           167
          8pm-9pm
                           131
          6pm-7pm
                            92
          7pm-8pm
                            61
          Past 10pm
                            54
          Past Midnight
                            38
         Name: Dinner time, dtype: int64
In [35]:
          label_encoder = LabelEncoder()
          df['Dinner time (Encoded)'] = label_encoder.fit_transform(df['Dinner time'])
          df[['Dinner time', 'Dinner time (Encoded)']]
```

```
Out[35]:
               Dinner time  Dinner time (Encoded)
            0
                  7pm-8pm
            1
                                              0
                 6pm-7pm
            2
                9pm-10pm
                                              3
            3
                  6pm-7pm
                                              0
                                              1
            4
                  7pm-8pm
          538
                  6pm-7pm
                                              0
          539
                9pm-10pm
                                              3
          540
                 7pm-8pm
                                              1
          541
                 Past 10pm
          542
                                              0
                 6pm-7pm
```

```
543 rows × 2 columns
```

Out[37]:		Variety	variety (Encoded)
	0	Italian	2
	1	Chinese	0
	2	Indian	1
	3	Chinese	0
	4	Chinese	0
	•••		
	538	Indian	1
	539	Chinese	0
	540	Indian	1
	541	Indian	1
	542	Indian	1

543 rows × 2 columns

```
In [38]: df
```

Out[38]:

	Register Number	Name	Date	Food Type	Location	Where in College ?	Variety	Avg money	Dinner time	Weekday
0	23112301	Arghya Basu	2024- 03-04	Veg	Market	Not in college	Italian	125.0	7pm- 8pm	Monday
1	23112302	Aryan Singhal	2024- 03-04	Non - Veg	College	Steaming Mugs	Chinese	225.0	6pm- 7pm	Monday
2	23112303	Avinash Ruthvik	2024- 03-04	Non - Veg	Market	Not in college	Indian	125.0	9pm- 10pm	Monday
3	23112305	Harshil Mishra	2024- 03-04	Non - Veg	Market	Not in college	Chinese	175.0	6pm- 7pm	Monday
4	23112306	Juhi Rathore	2024- 03-04	Veg	Home- Cooked	Not in college	Chinese	75.0	7pm- 8pm	Monday
•••										
538	23112316	Tarit Kumar Singh	2024- 04-15	Veg	College	Cafe by the Valley	Indian	125.0	6pm- 7pm	Monday
539	23112317	Urvi Saran	2024- 04-15	Veg	Market	Not in college	Chinese	175.0	9pm- 10pm	Monday
540	23112318	Veda Yogesh Rane	2024- 04-15	Veg	Market	Not in college	Indian	75.0	7pm- 8pm	Monday
541	23112320	Abhimanyu Sharma	2024- 04-15	Non - Veg	Market	Not in college	Indian	125.0	Past 10pm	Monday
542	23112321	Kratik Lodha	2024- 04-15	Non - Veg	Market	Not in college	Indian	225.0	6pm- 7pm	Monday

543 rows × 14 columns

```
In [39]: X = df[['Avg money (Normalized)', 'Dinner time (Encoded)','variety (Encoded)']]
         y = df['Where in College ?']
In [40]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_sta
         model2= LogisticRegression()
In [44]:
         model2.fit(X_train, y_train)
Out[44]:
         ▼ LogisticRegression
         LogisticRegression()
In [45]: y_pred = model2.predict(X_test)
         accuracy = accuracy_score(y_test, y_pred)
         print("Accuracy:", accuracy)
         Accuracy: 0.6697247706422018
         new_data = [[10, 20,0.3]]
In [46]:
         predicted_location = model2.predict(new_data)
         print("Predicted Location:", predicted_location)
```

Predicted Location: ['Cafe by the Valley']

```
In [47]:
    results_data = {
        'Actual_Location': y_test,
        'Predicted_Location': y_pred
}

results_df = pd.DataFrame(results_data)
results_df['Error'] = results_df['Actual_Location'] != results_df['Predicted_Location']
results_df.reset_index(drop=True, inplace=True)

results_df
```

```
Out[47]:
                   Actual_Location Predicted_Location
                                                           Error
               0
                      Not in college
                                           Not in college
                                                           False
              1
                         Chopsticks
                                           Not in college
                                                            True
                  Cafe by the Valley
                                           Not in college
                                                            True
              3
                      Not in college
                                           Not in college
                                                           False
               4
                      Not in college
                                           Not in college
                                                            False
            104
                      Not in college
                                           Not in college
                                                           False
            105
                      Not in college
                                           Not in college
                                                           False
            106 Cafe by the Valley
                                           Not in college
                                                            True
            107
                      Not in college
                                           Not in college
                                                           False
            108
                      Not in college
                                           Not in college False
           109 rows × 3 columns
```

```
In [48]: new_df = df.drop(columns=['Register Number'])
    new_df
```

Out[48]:

	Name	Date	Food Type	Location	Where in College ?	Variety	Avg money	Dinner time	Weekday	Avg mond (Normalized
0	Arghya Basu	2024- 03-04	Veg	Market	Not in college	Italian	125.0	7pm- 8pm	Monday	0.3333
1	Aryan Singhal	2024- 03-04	Non - Veg	College	Steaming Mugs	Chinese	225.0	6pm- 7pm	Monday	1.00000
2	Avinash Ruthvik	2024- 03-04	Non - Veg	Market	Not in college	Indian	125.0	9pm- 10pm	Monday	0.3333:
3	Harshil Mishra	2024- 03-04	Non - Veg	Market	Not in college	Chinese	175.0	6pm- 7pm	Monday	0.66666
4	Juhi Rathore	2024- 03-04	Veg	Home- Cooked	Not in college	Chinese	75.0	7pm- 8pm	Monday	0.00000
•••							•••			
538	Tarit Kumar Singh	2024- 04-15	Veg	College	Cafe by the Valley	Indian	125.0	6pm- 7pm	Monday	0.3333:
539	Urvi Saran	2024- 04-15	Veg	Market	Not in college	Chinese	175.0	9pm- 10pm	Monday	0.66666
540	Veda Yogesh Rane	2024- 04-15	Veg	Market	Not in college	Indian	75.0	7pm- 8pm	Monday	0.00000
541	Abhimanyu Sharma	2024- 04-15	Non - Veg	Market	Not in college	Indian	125.0	Past 10pm	Monday	0.3333
542	Kratik Lodha	2024- 04-15	Non - Veg	Market	Not in college	Indian	225.0	6pm- 7pm	Monday	1.00000

543 rows × 13 columns

In [49]: sns.heatmap(new\_df.corr(),annot=True,cmap="viridis")

Out[49]: <Axes: >

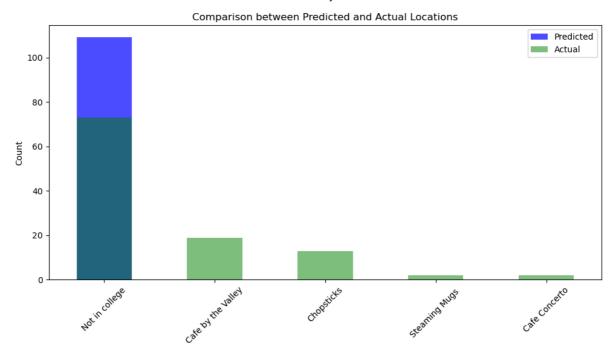


In [50]: df.corr()

Out[50]:

	Register Number	Avg money	Avg money (Normalized)	Dinner time (Encoded)	variety (Encoded)
Register Number	1.000000	0.008144	0.008144	0.032416	0.026840
Avg money	0.008144	1.000000	1.000000	0.004923	-0.030504
Avg money (Normalized)	0.008144	1.000000	1.000000	0.004923	-0.030504
Dinner time (Encoded)	0.032416	0.004923	0.004923	1.000000	0.002721
variety (Encoded)	0.026840	-0.030504	-0.030504	0.002721	1.000000

```
In [51]: plt.figure(figsize=(10, 6))
    results_df['Predicted_Location'].value_counts().plot(kind='bar', color='blue', alph
    results_df['Actual_Location'].value_counts().plot(kind='bar', color='green', alpha=
    plt.title('Comparison between Predicted and Actual Locations')
    plt.xlabel('Location')
    plt.ylabel('Count')
    plt.legend()
    plt.xticks(rotation=45)
    plt.tight_layout()
    plt.show()
```



Location

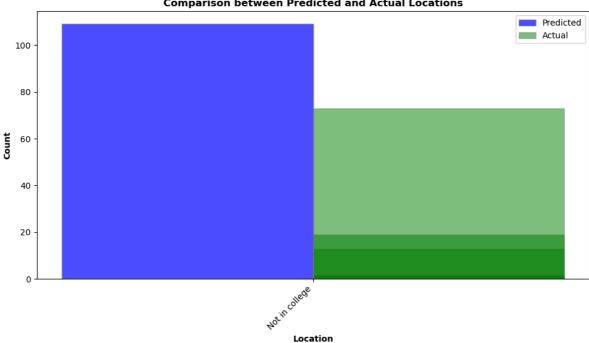
```
In [52]: fig, ax = plt.subplots(figsize=(10, 6))
    locations = results_df['Predicted_Location'].unique()
    bar_width = 0.35

    r1 = range(len(locations))
    r2 = [x + bar_width for x in r1]

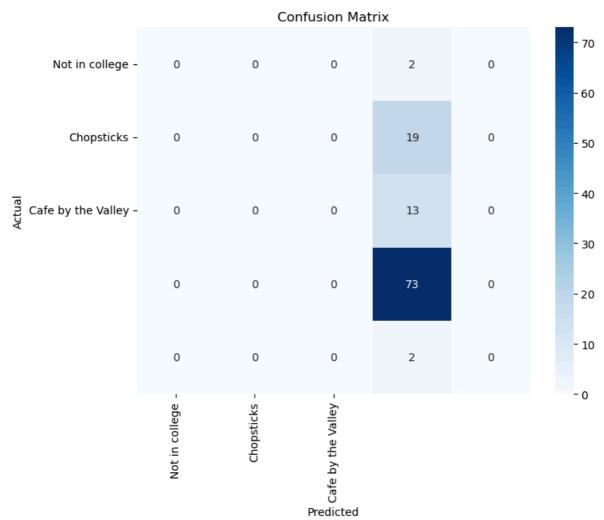
    predicted_bars = ax.bar(r1, results_df['Predicted_Location'].value_counts().sort_ir actual_bars = ax.bar(r2, results_df['Actual_Location'].value_counts().sort_index(),
    ax.set_xlabel('Location', fontweight='bold')
    ax.set_ylabel('Count', fontweight='bold')
    ax.set_title('Comparison between Predicted and Actual Locations', fontweight='bold' ax.set_xticks([r + bar_width/2 for r in range(len(locations))])
    ax.set_xticklabels(locations, rotation=45, ha='right')
    ax.legend()

plt.tight_layout()
    plt.show()
```

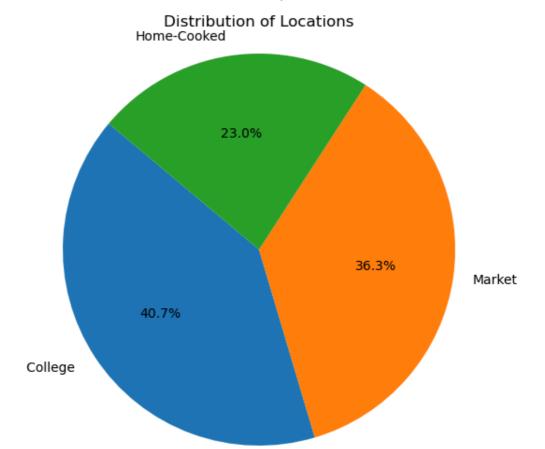




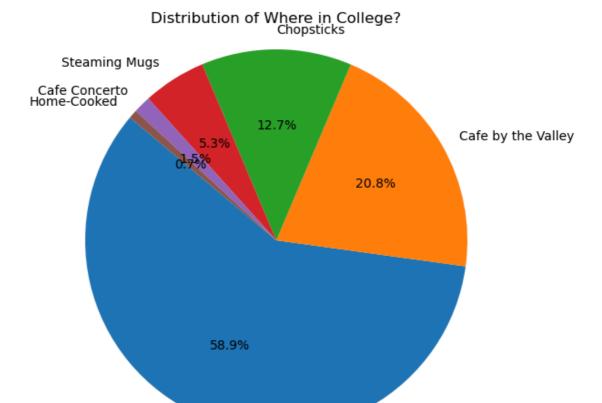
```
In [ ]:
In [53]:
         cm = confusion_matrix(results_df['Actual_Location'], results_df['Predicted_Location']
         plt.figure(figsize=(8, 6))
         sns.heatmap(cm, annot=True, cmap='Blues', fmt='g', xticklabels=['Not in college',
         plt.xlabel('Predicted')
         plt.ylabel('Actual')
         plt.title('Confusion Matrix')
         plt.show()
```



```
In [54]: location_counts = df['Location'].value_counts()
    plt.figure(figsize=(8, 6))
    plt.pie(location_counts, labels=location_counts.index, autopct='%1.1f%%', startangl
    plt.title('Distribution of Locations')
    plt.axis('equal')
    plt.show()
```



```
In [55]: where_counts = df['Where in College ?'].value_counts()
   plt.figure(figsize=(8, 6))
   plt.pie(where_counts, labels=where_counts.index, autopct='%1.1f%%', startangle=140)
   plt.title('Distribution of Where in College?')
   plt.axis('equal')
   plt.show()
```



In [56]: **df** 

Not in college

Out[56]:

Where Register Food in Avg Dinner Date Variety Name Location Weekday Number Type College money time ? Arghya 2024-Not in 7pm-0 23112301 Veg Market Italian 125.0 Monday Basu 03-04 college 8pm Aryan 2024-Non Steaming 6pm-23112302 225.0 College Chinese Monday Singhal 03-04 - Veg Mugs 7pm Avinash 2024-Non Not in 9pm-2 23112303 Market Indian 125.0 Monday Ruthvik 03-04 - Veg college 10pm Harshil 2024-Non Not in 6pm-23112305 Market Chinese 175.0 Monday Mishra 03-04 college 7pm - Veg Juhi 2024-Home-Not in 7pm-75.0 23112306 Veg Chinese Monday Rathore 03-04 Cooked college 8pm Cafe by Tarit Kumar 2024-6pm-23112316 Veg College the Indian 125.0 Monday Singh 04-15 7pm Valley 2024-Not in 9pm-23112317 175.0 Monday 539 Urvi Saran Market Chinese Veg 04-15 college 10pm Veda 2024-Not in 7pm-540 23112318 Market Indian 75.0 Monday Yogesh Veg 04-15 college 8pm Rane Abhimanyu 2024-Non Not in **Past 541** 23112320 Market Indian 125.0 Monday Sharma 10pm 04-15 - Veg college Kratik 2024-Non Not in 6pm-**542** 23112321 225.0 Market Indian Monday Lodha 04-15 - Veg college 7pm

543 rows × 14 columns

Average spent: 75.0 What they ate: Indian

Enter the date (yyyy-mm-dd): 2024-04-04

Where they ate: Not in college

```
y_bin = label_binarize(results_df['Actual_Location'], classes=['Not in college', '(
In [58]:
         fpr = dict()
         tpr = dict()
         roc_auc = dict()
         for i in range(len(y_bin[0])):
             fpr[i], tpr[i], _ = roc_curve(y_bin[:, i], results_df['Predicted_Location'] ==
              roc_auc[i] = auc(fpr[i], tpr[i])
         plt.figure(figsize=(8, 6))
         colors = ['blue', 'red', 'green']
         for i, color in zip(range(len(y_bin[0])), colors):
             plt.plot(fpr[i], tpr[i], color=color, lw=2, label='ROC curve of class {0} (area
         plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
         plt.xlim([0.0, 1.0])
         plt.ylim([0.0, 1.05])
         plt.xlabel('False Positive Rate')
         plt.ylabel('True Positive Rate')
         plt.title('Receiver Operating Characteristic (ROC) Curve')
         plt.legend(loc="lower right")
         plt.show()
```

## Receiver Operating Characteristic (ROC) Curve 1.0 0.8 **Frue Positive Rate** 0.6 0.4 0.2 ROC curve of class 0 (area = 0.50) ROC curve of class 1 (area = 0.50) ROC curve of class 2 (area = 0.50) 0.0 0.2 0.4 0.0 0.6 0.8 1.0 False Positive Rate

```
import pickle
with open('Dinner prediction.pkl', 'wb') as file:
    pickle.dump(model2, file)

In []:
from flask import Flask, request, jsonify
from flask_cors import CORS
import pickle
import pandas as pd

app = Flask(__name__)
CORS(app,support_credentials=True)
```

```
with open('Dinner prediction.pkl', 'rb') as model_file:
            model = pickle.load(model_file)
        @app.route("/sampleGet")
        def sample_endpoint():
            return "Yes, Server Received!"
        @app.route('/predict_location', methods=['POST'])
        def predict_location():
            data = request.json
            new_data = [[data['avg_money'], data['dinner_time'], data['variety']]]
            predicted_location = model.predict(new_data)
            return jsonify({"predicted_location": predicted_location[0]})
        @app.route('/get_info', methods=['POST'])
        def get_info():
            data = request.json
            register_number = data['register_number']
            date = data['date']
            filtered_df = df[(df['Register Number'] == register_number) & (df['Date'] == data
            if filtered df.empty:
                 return jsonify({"message": "No information found for the given register num
            where_ate = filtered_df.iloc[0]['Where in College ?']
            avg_spent = filtered_df['Avg money (Normalized)'].mean()
            what_ate = filtered_df.iloc[0]['Variety']
            return jsonify({
                 "where_ate": where_ate,
                 "avg_spent": avg_spent,
                 "what_ate": what_ate
            })
        if __name__ == '__main__':
            app.run(host='0.0.0.0',port='8000')
         * Serving Flask app '__main__'
         * Debug mode: off
        WARNING: This is a development server. Do not use it in a production deployment. U
        se a production WSGI server instead.
         * Running on all addresses (0.0.0.0)
         * Running on http://127.0.0.1:8000
         * Running on http://10.21.104.113:8000
        Press CTRL+C to quit
        10.21.104.113 - - [15/May/2024 13:10:37] "OPTIONS /predict_location HTTP/1.1" 200
        10.21.104.113 - - [15/May/2024 13:10:37] "POST /predict location HTTP/1.1" 200 -
In [ ]:
In [ ]:
```