Part 2 - Restaurant Project 2 for MGMT 635 (Data Mining for Managers)
(Swasti Johri)

Abstract: The restaurant industry has experienced significant transformation due to the COVID-19 pandemic, leading many businesses to reassess their profit-enhancing strategies. Using a dataset that includes various operational factors from several Tri-state area restaurants, my analysis was aimed at pinpointing the best strategies for revenue increase. I utilized both a linear regression and a neural network model to recognize determinants that could forecast expected revenues. The results of my study indicated that the existence of a restaurant app had no notable effect on revenue. On the contrary, the area's income and the population of the town were the two major contributors to revenue. Data also showed a positive connection between revenue and take-out services, while a negative correlation was observed with outdoor dining services. The neural network model gave more promising revenue predictions compared to the linear regression model, especially for restaurants TRE and WRT, which showed the highest potential revenue growth. However, restaurant GFR was an anomaly with a difference of roughly 1,611,176. Based on these findings, my suggestion to new restaurants is to concentrate on offering take-out services to attract local customers within a 2–3-mile distance from the town center and give preference to outdoor services rather than investing in app networks to best utilize their resources.

Introduction: The COVID-19 pandemic has presented substantial hurdles to the restaurant industry, as many establishments grapple with adapting to new modes of operation. Consequently, I've scrutinized the dataset containing various operational factors from a multitude of Tri-state area restaurants to discern the most effective strategies for enhancing revenue. The aim of this study is to uncover elements that can aid in forecasting anticipated revenues and pinpoint the most efficient methods to augment revenue. The following questions are to be addressed:

1. What are the revenue patterns among different restaurants? (Restaurants, Monthly Revenue)
2. Does technology have an impact on the revenue of various restaurants? (Restaurants, App, Monthly Revenue)
3. Are the revenue trends influenced by socio-economic factors? (Restaurant, Town Population, Area Income, Monthly Revenue)
4. Are consumers willing to commute from downtown for an outdoor dining experience? (Outdoor dining, Distance from town center, Monthly Revenue, Restaurants)
5. Is there a connection between take-out service and town population, which in turn affects monthly revenue?

Methodology:

Correlation Matrix-

	App	Take Out Service	Area Income	Town Population	Outdoor Dining	Distance from Town Center	Monthly Revenue
App	1	-0.01	0.09	0.12	0.15	-0.01	0.25
Take Out Service	-0.13	1	-0.08	0.07	-0.15	0.05	0.39
Area Income	0.11	-0.09	1	-0.16	0.06	-0.14	0.59
Town Population	0.18	0.06	-0.16	1	0.13	0.13	0.32
Outdoor Dining	0.17	-0.13	0.08	0.14	1	-0.16	0.47
Distance from Town Center	-0.11	0.06	-0.15	0.17	-0.16	1	-0.06
Monthly Revenue	0.25	0.39	0.59	0.32	0.47	-0.06	1

To fulfill my objectives, I formulated both a linear regression and neural network model to pinpoint factors that could potentially forecast expected revenues. I took into consideration various operational elements such as the presence of an app network, the income of the area, the town's population, the availability of take-out services, and outdoor dining options.

Looking at the correlation matrix, it seems the variables most significantly linked with monthly revenue are outdoor dining (0.47), area income (0.59), and take-out service (0.39).

These correlations could serve as an initial guide in planning a restaurant setup that is poised to maximize revenue.
Keeping these findings in mind, here are a few suggestions for constructing a restaurant that has the potential to maximize revenue:

1. Prioritize outdoor dining: Among the variables in the dataset, this one showed the strongest association with monthly revenue. Establishments that provide outdoor dining may be able to charge more for the unique experience, especially in high-income locales. If outdoor dining isn't feasible, think about integrating features that give the dining area an open, spacious feel.

2. Choose a high-income location: Restaurants situated in high-income areas tend to bring in more revenue. It's crucial, though, to weigh this against other aspects, such as the rent costs in the vicinity.

3. Provide take-out service: Eateries offering take-out service also seem to generate higher revenue. This could be especially beneficial in areas with less population density or where parking is a challenge, as customers might prefer to opt for takeout rather than dine in.

4. Contemplate the proximity to the town center: Although the dataset doesn't show a strong link between distance from the town center and revenue, it's still a factor worth considering. Restaurants closer to the town center might benefit from higher visibility and accessibility.

I employed a Multi-Layer Perceptron (a type of neural network model) to calculate the outcomes, which resulted in predictions accompanied by a substantial amount of test error values.

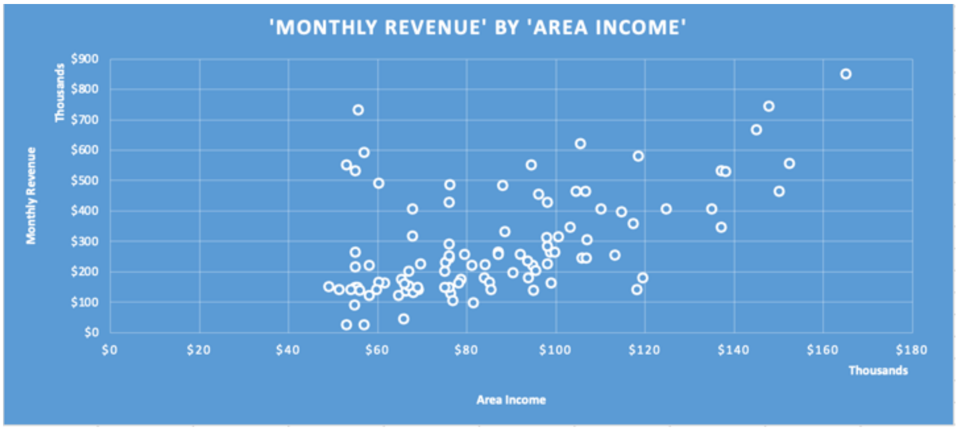
```
In [43]: predictions = model.predict(new_data)

         1/1 [==============================] - 0s 47ms/step

In [44]: print(predictions)

         [[146481.1 ]
          [325826.12]
          [339064.06]
          [285568.88]
          [254641.25]
          [246984.22]
          [345088.88]
          [182735.92]
          [133649.12]
          [223420.69]
          [239498.38]]
```

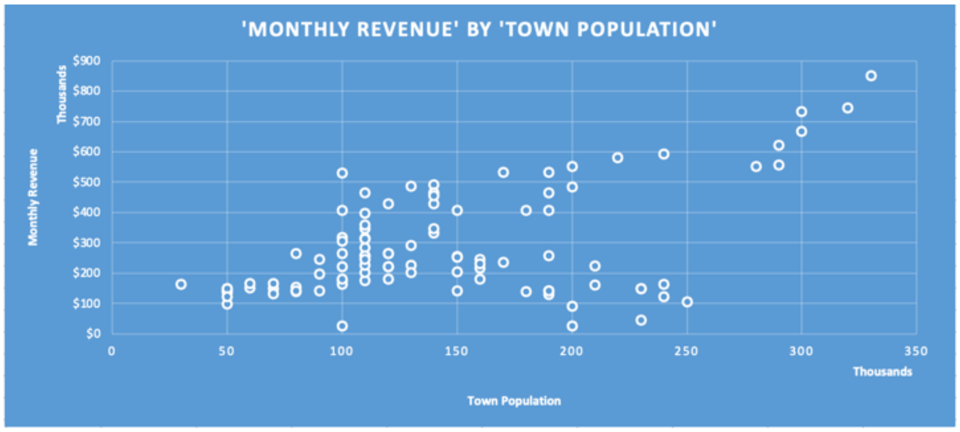
Therefore, I improved the model's performance using the Grid Search Method from Keras, in combination with Multiple Linear Regression. Please refer to the subsequent code for details-

```
# Multiple Linear Regression
from sklearn import linear_model
regr = linear_model.LinearRegression()
regr.fit(x_train,y_train)
MLR_train_pred = regr.predict(x_train)
MLR_test_pred = regr.predict(x_test)
coefficients = regr.coef_
df_coef = pd.DataFrame(x_train.columns)
df_coef["Coefficient"] = coefficients.tolist()

MLR_ME_train = round(abs(np.subtract(y_train,MLR_train_pred)).mean(),2)
MLR_MSE_train = round(np.square(np.subtract(y_train,MLR_train_pred)).mean(),2)
print("Following are the results for the Multiple Linear Regression Algorithm")
print("The mean error for the train set is "+str(MLR_ME_train))
print("The mean squared error for the train set is "+str(MLR_MSE_train))


# Neural Network SKLearn
from sklearn.model_selection import GridSearchCV
from sklearn.neural_network import MLPRegressor

#Suppress warnings when applying GridSearchCV to find best Neural Network constants
from warnings import simplefilter
from sklearn.exceptions import ConvergenceWarning
simplefilter("ignore", category=ConvergenceWarning)


param_list = {"hidden_layer_sizes": [(1,),(50,)], "activation": ["identity", "logistic", "tanh", "relu"], "solver": ["lbfgs", "sgd", "adam"], "alpha": [0.00005,0.0005]}

clf = GridSearchCV(MLPRegressor(max_iter=1000),param_list, n_jobs=-1, cv=3)
clf = clf.fit(x_train, y_train)
NN_test_pred = clf.predict(x_test)
NN_train_pred = clf.predict(x_train)

NN_ME_train = round(abs(np.subtract(y_train,NN_train_pred)).mean(),2)
NN_MSE_train = round(np.square(np.subtract(y_train,NN_train_pred)).mean(),2)

print("Following are the results for the SKLearn Neural Network")
print("The mean error for the train set is "+str(NN_ME_train))
print("The mean squared error for the train set is "+str(NN_MSE_train))
```

Data Purification – Preprocessing

```
import pandas as pd
import numpy as np


sheet_url = "https://docs.google.com/spreadsheets/d/1lIdYmIr-GRxdYdohfO01i1PUrNoAk_AB/edit#gid=1233047613"
url = sheet_url.replace("/edit#gid=", "/export?format=csv&gid=")
df = pd.read_csv(url)
df.drop(columns=df.columns[-9:],axis=1, inplace=True)
df_before = df


sheet_url = "https://docs.google.com/spreadsheets/d/1lIdYmIr-GRxdYdohfO01i1PUrNoAk_AB/edit#gid=100581497"
url = sheet_url.replace("/edit#gid=", "/export?format=csv&gid=")
df_2 = pd.read_csv(url)
df = df.append(df_2, ignore_index = True)
df_before2 = df

columns = ["Area Income", "Monthly Revenue", "Town Population"]
for i in columns:
    df[i] = df[i].str.replace("$","", regex=True).str.replace(",","", regex=True)
    df[i] = df[i].apply(pd.to_numeric)
one_hot = pd.get_dummies(df["Restaurants"])
df = df.join(one_hot)
df.drop(columns="Restaurants", inplace=True)

#Option to take out Restaurant Encoding

df.drop(columns=one_hot.columns, inplace=True)


# Train & Test Split
y = df["Monthly Revenue"]
x = df.drop(columns="Monthly Revenue")

# Standardize Values (0-1)
for i in x.columns:
    x[i] = x[i]/x[i].max()

y_train = y[0:-11]
x_train = x[0:-11]
x_test = x[-11:]
```





The analysis suggests that area income and town population are the two most influential factors impacting revenues.

The graphical representation reveals that the relationship between area income and revenue is denser and non-linear, whereas the connection between town population and revenue seems to be more dispersed and linear.

Based on these findings, it can be inferred that the population of a town could be a more significant contributor to revenue generation than the income of the area.

Upon reviewing the data, it was determined that two elements had a negative influence on the monthly revenues of the restaurants: the establishment's distance from the town center and the presence of an app network. Thus, it is advised that restaurants focus on securing an optimal town location over investing resources in an app network.

Four factors have been identified as having a strong positive relationship with the monthly revenue: Take-Out Services, Town Population, Area Income, and Outdoor Services. Among these, the influence of Town Population and Area Income on the revenue is slightly more pronounced than the others. Notably, the town's population has a greater bearing on the revenue. As a result, I recommend that restaurants' concentrate on offering take-out and outdoor services to attract a larger customer base, particularly in regions with a denser population and a higher income bracket.

	Restaurants	SKLearn Neural Network	SKLearn Multiple Linear Regression
99	PPI	255930.417353	122614.748603
100	TRE	665520.486910	407901.781564
101	GGT	357012.883496	377433.713612
102	MND	503255.286104	328495.265542
103	WRT	592090.753434	543535.924556
104	GFR	309634.704639	192810.337631
105	WWW	536580.836351	404494.604147
106	QWE	194528.685287	239704.369952
107	FGR	120712.566626	152314.485707
108	SSC	326563.317405	107684.757304
109	SAE	237610.856464	291572.418225

Findings: My study revealed that an app network's existence didn't significantly impact the revenue, while the income of the area and the town's population emerged as the two most crucial determinants of revenue. There was a positive relationship between take-out services and revenue, whereas outdoor dining services were negatively correlated.

The revenue forecasts from my neural network model were more promising compared to those from the linear regression model, with TRE and WRT restaurants showing the highest projected revenues. Restaurant GFR, however, was an anomaly, exhibiting a revenue difference of approximately 1,611,176.

Strategy Recommendation: Drawing from these observations, I advise new restaurants to prioritize providing take-out services to appeal to customers within a 2–3-mile radius from the town center and rank outdoor services higher than app networks to best allocate their resources.

Furthermore, I recommend that restaurant proprietors' factor in the area's socio-economic aspects when choosing their establishment's location. Locations with larger town populations are more likely to result in higher revenues compared to areas with high income. It's crucial to contemplate the expenses tied to setting up, maintaining, and repairing outdoor dining services as these could impact the profit margins.

Closing Remarks: My examination of the dataset encompassing various operational factors from numerous Tri-state area restaurants has shed light on how to tweak restaurant operations to boost revenues. The data suggests that new restaurants should emphasize take-out services, position themselves within a 2–3-mile radius of the town center, and assign higher priority to outdoor services over app networks. Understanding the area's socio-economic aspects can aid owners in picking the most advantageous location for their establishment.

By integrating these tactics and variables into their business strategy, new restaurants can maximize their revenue potential and enhance their overall business performance.