Project Report: Evaluating the Effectiveness of Marketing Strategies on Sales of a New Menu Item.

Objective:

The primary aim of this analysis is to evaluate the impact of three different promotions(marketing strategies) on sales of a new menu item at a fast-food chain and determine the most effective strategy.

DatasetOverview:

- Initial Dataset Shape:The dataset consists of 459 records and 7 columns.
  - MarketID: unique identifier for market
  - MarketSize: size of market area by sales
  - LocationID: unique identifier for store location
  - AgeOfStore: Age of store in years
  - Promotion: one of three promotions that were tested
  - week: one of four weeks when the promotions were run
  - SalesInThousands: sales amount for a specific LocationID, Promotion, and week.

```
products = pd.read_csv('/content/WA_Marketing-Campaign.csv')
products.head()
```

	MarketID	MarketSize	LocationID	AgeOfStore	Promotion	week	SalesInThousands
0	1	Medium	1	4	3	1	33.73
1	1	Medium	1	4	3	2	35.67
2	1	Medium	1	4	3	3	29.03
3	1	Medium	1	4	3	4	39.25
4	1	Medium	2	5	2	1	27.81

Tools and Libraries :

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd
```

Checking for the null values in data

```
products.isnull().sum()
```

	0
MarketID	0
MarketSize	0
LocationID	0
AgeOfStore	0
Promotion	0
week	0
SalesInThousands	0

dtype: int64

The data doesn't have null values. Now let's have a look at the column insights before moving forward.

```
products.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 548 entries, 0 to 547
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   MarketID          548 non-null    int64
 1   MarketSize        548 non-null    object
 2   LocationID        548 non-null    int64
 3   AgeOfStore        548 non-null    int64
 4   Promotion         548 non-null    int64
 5   week              548 non-null    int64
 6   SalesInThousands  548 non-null    float64
dtypes: float64(1), int64(5), object(1)
memory usage: 30.1+ KB
```

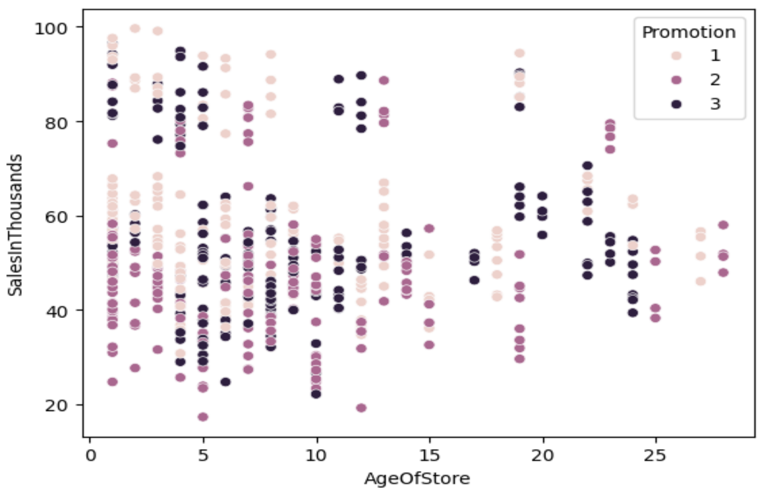
Now let's have a look at the descriptive statistics of the data:

```
products.describe()
```

	MarketID	LocationID	AgeOfStore	Promotion	week	SalesInThousands
count	548.000000	548.000000	548.000000	548.000000	548.000000	548.000000
mean	5.715328	479.656934	8.503650	2.029197	2.500000	53.466204
std	2.877001	287.973679	6.638345	0.810729	1.119055	16.755216
min	1.000000	1.000000	1.000000	1.000000	1.000000	17.340000
25%	3.000000	216.000000	4.000000	1.000000	1.750000	42.545000
50%	6.000000	504.000000	7.000000	2.000000	2.500000	50.200000
75%	8.000000	708.000000	12.000000	3.000000	3.250000	60.477500
max	10.000000	920.000000	28.000000	3.000000	4.000000	99.650000

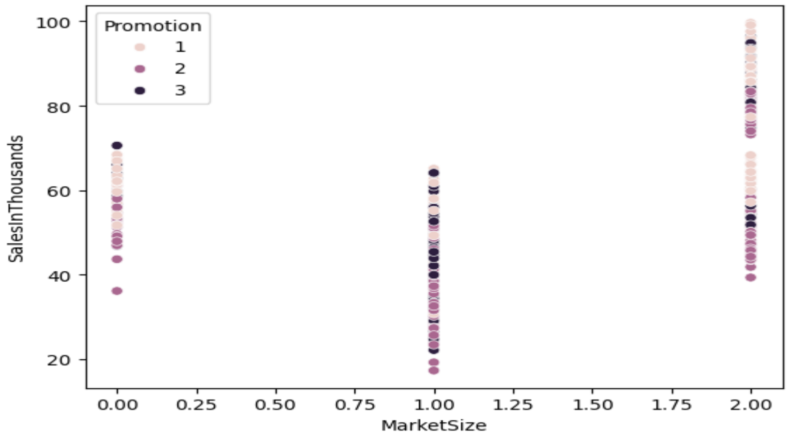
Let's have a look at the relationship between AgeofStore and SalesInThousands for all three promotions:

```
sns.scatterplot(data = products,x = 'AgeOfStore',y
='SalesInThousands' ,hue = 'Promotion')
```



Interpretation: The relationship between the age of the store and sales remains consistent across all promotions, as shown by the scatterplot. This indicates that whether a store is old or newly opened, sales are the same for all promotions.

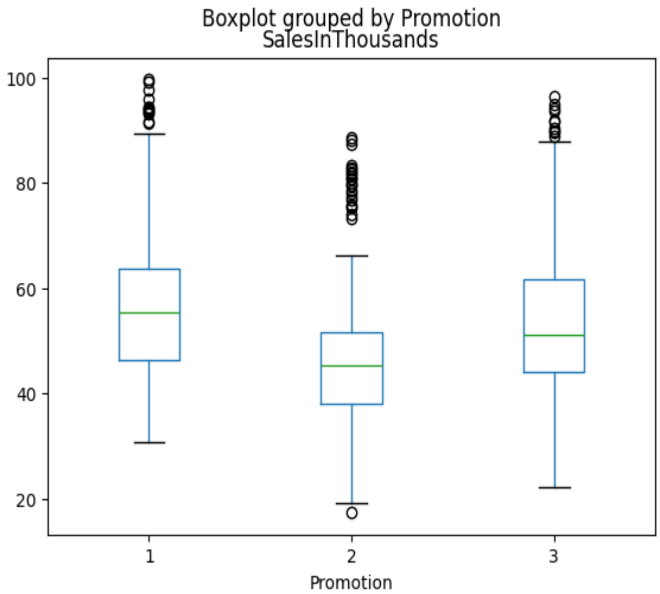
```
products['MarketSize'] =
products['MarketSize'].map({'Small':0,'Medium':1,'Large':2})
sns.scatterplot(data = products,x = 'MarketSize',y =
'SalesInThousands',hue = 'Promotion')
```



Interpretation: The relationship between the Marketsize and sales remains consistent across all promotions, as shown by the scatterplot. This indicates that whether a store is Small, Medium and Large sales are the same for all promotions.

Checking For Outliers:

```
products.boxplot(column = 'SalesInThousands',by ='Promotion',grid =
False )
```



Interpretation: As we can see from the boxplot, there are outliers in the data. To handle them, I used the capping method.

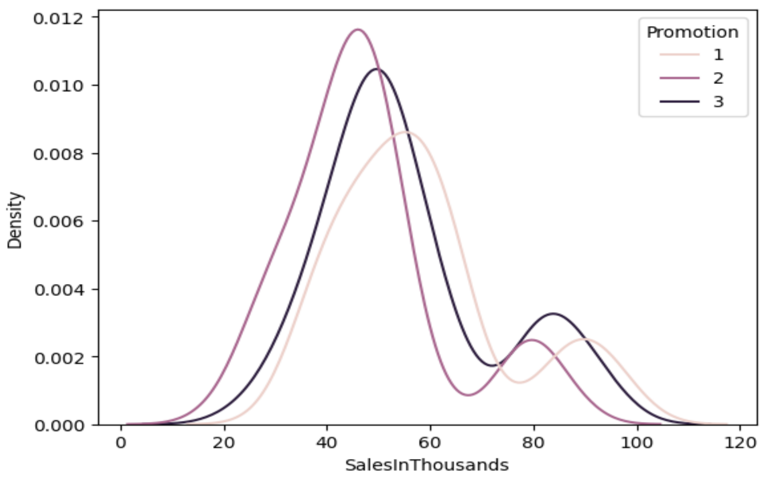
```
Q21=
products[products['Promotion']==1]['SalesInThousands'].quantile(0.50
)
print(Q21)
var1= products[products['Promotion']==1]['SalesInThousands'].std()
upper_limit_1 = Q21+2*var1
lower_limit_1 = Q21-2*var1
```

```
print(upper_limit_1,lower_limit_1)
products.loc[products['Promotion']==1,'SalesInThousands'] =
products.loc[products['Promotion']==1,'SalesInThousands'].clip(lower
_limit_1,upper_limit_1)

Q22=
 products[products['Promotion']==2]['SalesInThousands'].quantile(0.5
0)
print(Q22)
var2= products[products['Promotion']==2]['SalesInThousands'].std()
upper_limit_2 = Q22+2*var2
lower_limit_2 = Q22-2*var2
print(upper_limit_2,lower_limit_2)
products.loc[products['Promotion']==2,'SalesInThousands'] =
products.loc[products['Promotion']==2,'SalesInThousands'].clip(lower
_limit_2,upper_limit_2)

Q23=
 products[products['Promotion']==2]['SalesInThousands'].quantile(0.5
0)
print(Q23)
var3= products[products['Promotion']==2]['SalesInThousands'].std()
upper_limit_3 = Q23+2*var3
lower_limit_3 = Q23-2*var3
print(upper_limit_3,lower_limit_3)
products.loc[products['Promotion']==3,'SalesInThousands'] =
products.loc[products['Promotion']==3,'SalesInThousands'].clip(lower
_limit_3,upper_limit_3)

products.boxplot(column = 'SalesInThousands',by ='Promotion',grid =
False )
```

Now, we perform ANOVA testing To check the impact of promotions on sales. For that we firstly check assumptions of ANOVA.

1.Check Normality of data

```
sns.kdeplot(data = products,x = 'SalesInThousands',hue =
'Promotion')
```



Interpretation: As observed from the density plots of the Promotions data, the distribution is not perfectly normal. However, since our sample size is greater than 30, the Central Limit Theorem allows us to proceed with ANOVA.

2. Homogeneity of Variance : We check this assumption of ANOVA by using Levene's Test:

```
stat, p = levene(products[products['Promotion'] ==
1]['SalesInThousands'],
                 products[products['Promotion'] ==
2]['SalesInThousands'],
                 products[products['Promotion'] ==
3]['SalesInThousands'])
print(p-value:p)
```

```
p-value:0.2817
```

Interpretation: As we observe, the p-value of Levene's test is greater than the significance level ($\alpha = 0.05$), so we fail to reject the null hypothesis. This indicates that the variances of the data for the three promotions are equal.

Both assumptions of ANOVA are satisfied, so we can proceed.

Now, we perform ANOVA

```
f_stat,p_value =
f_oneway(products[products['Promotion']==1]['SalesInThousands'],

products[products['Promotion']==2]['SalesInThousands'],

products[products['Promotion']==3]['SalesInThousands'])

print("f_statistics":f_stat,"p_value":p_value)
```

```
f_statistics:27.7108 p_value:3.4567e-12
```

Interpretation: In comparing the sales of new food based on promotions, we conducted an ANOVA test to determine whether there is a statistically significant difference in sales among the three promotions. The results of the ANOVA test are as follows:

- **Test statistic (F-statistic):** 27.7108
- **p-value:** 3.4567×10^{-12}

The null hypothesis assumes that there is no statistically significant difference in sales among the three promotions. In this case, the p-value is approximately 3.4567×10^{-12}, which is much smaller than the typical significance level of 0.05.

- Since the p-value is less than 0.05, we reject the null hypothesis. This indicates that there is a statistically significant difference in sales among the three promotions.

Now,we have to find best strategy among these three for that we use Tukey's HSD test

```python
tuky = pairwise_tukeyhsd(endog = products['SalesInThousands'],groups
= products['Promotion'],alpha = 0.05)
print(tuky)
```

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2 meandiff p-adj   lower    upper  reject
-----------------------------------------------------
     1      2 -10.8933    0.0 -14.3722 -7.4143   True
     1      3  -4.5653 0.0061  -8.0442 -1.0863   True
     2      3    6.328    0.0   2.9273  9.7288   True
-----------------------------------------------------
```

Interpretation: From the Tukey test results, we observe that **Promotion 2** is the most effective in increasing sales, as it has significantly higher mean sales compared to both Promotion 1 and Promotion 3

Conclusion & Recommendations:

"Based on these results, it is recommended to implement the second marketing strategy to maximize the sales of the new menu item."