Advanced Assignment: Spreadsheet & Statistics

1. Business Understanding - Advanced

Background

- Fast-food chain plans to add a new item to its menu.
- The company is still undecided between three possible marketing campaigns for promoting the new product.
- In order to determine which promotion has the greatest effect on sales, the new item is introduced at locations in several randomly selected markets.

1. Business Understanding - Advanced

Core Business Problem

- Company wants to know which among 3 marketing campaigns yield the best sales.
- Company wants to know whether market size has effect on sales.

1. Business Understanding - Advanced

How data provided could help answer business problem

The dataset provides informations about sales performance of each marketing campaigns so that I can run statistical test to infer the conclusion into real population.

1. Business Understanding - Advanced

Statistics that can answer business problems

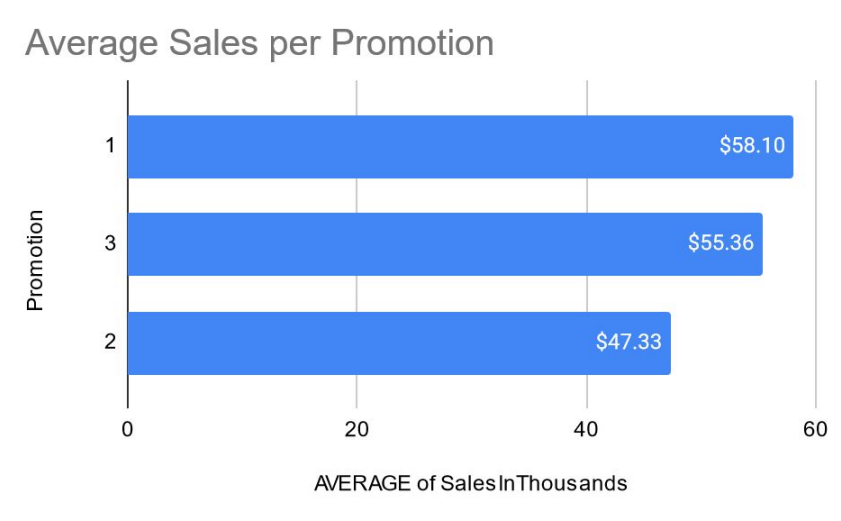
- After cleaning the data, exploratory data analysis are the first thing I can do.
- To generalize insight from sample to population, I need to run hypothesis test.

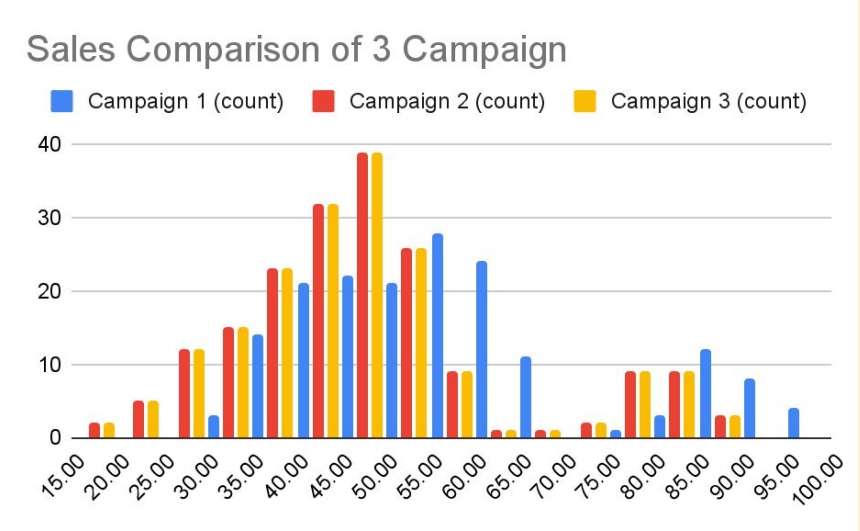
For example:

Correlation study or T-test

2. Exploratory Data Analysis

Marketing Campaign Performance





INSIGHTS:

- Marketing Campaign 1 have better average sales than other.
- All 3 marketing campaign have double peak in their distribution, indicating there are other factor that influences sales other than marketing campaign.

GSHEET LINK

2. Exploratory Data Analysis

Character of Store Based on Market Size

	Small	Medium	Large
Number of Unique Store	15	80	42
Average age	10.8	8.8	7.1
Average Sales	$57,409	$43,985	$70,116
Sales Sum	$3408.39	$14000.06	$11779.61

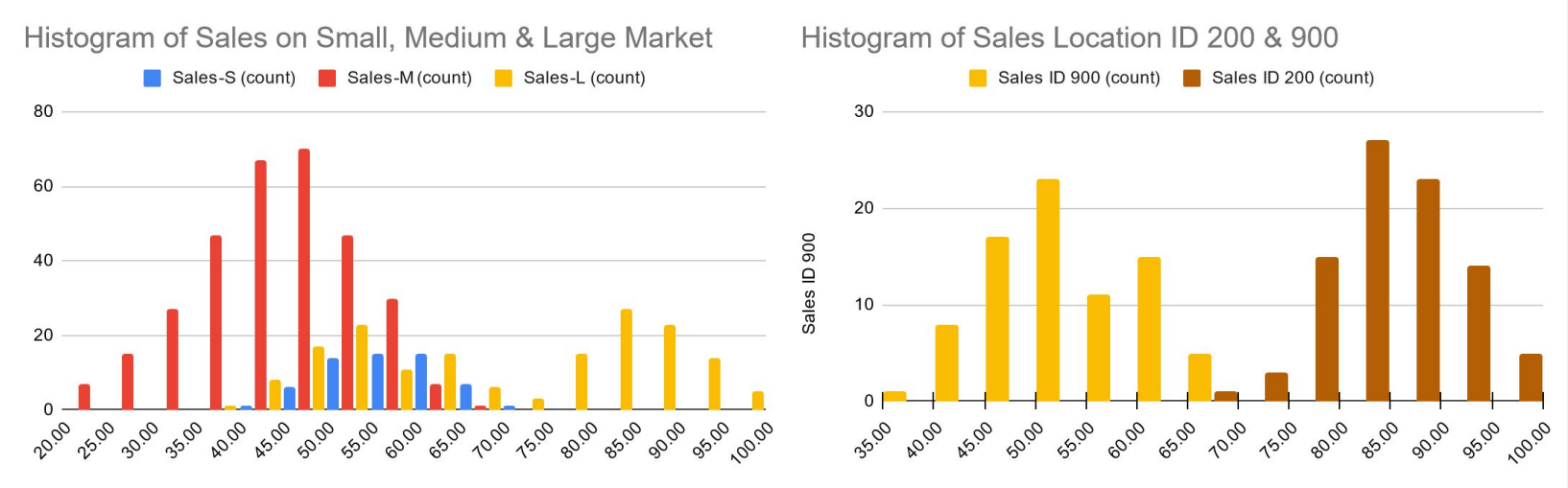
INSIGHTS:

- Medium market size store is the most represented sample and the small size store is the least represented.
- Larger market size store tend to have lesser average age.
- In the sample, medium market size stores generate the most sales, although the average sales per store is the lowest.

GSHEET LINK

2. Exploratory Data Analysis

Character of Store Based on Market Size



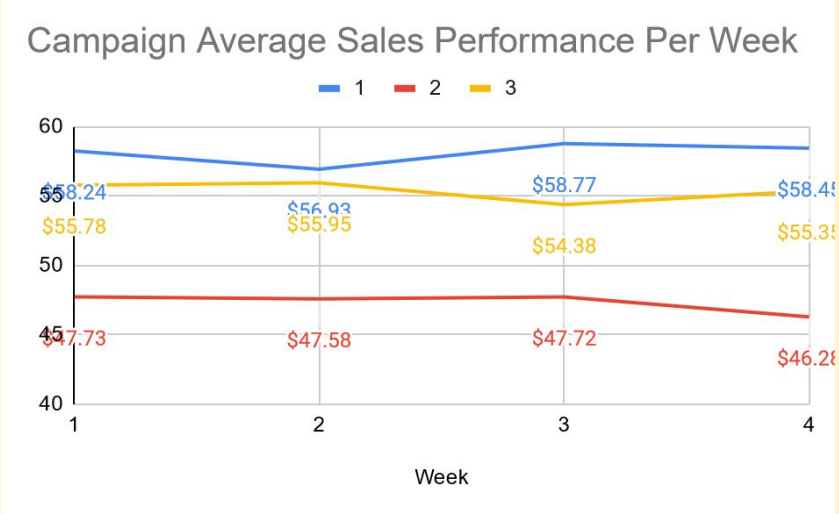
INSIGHTS:

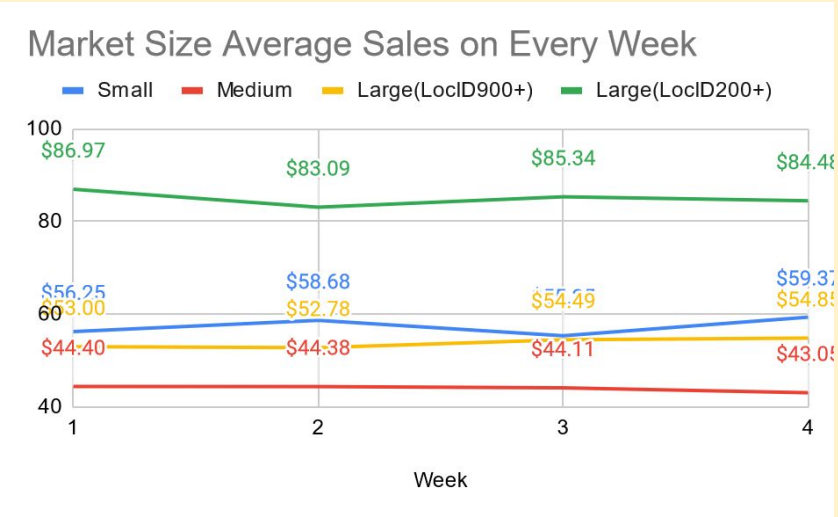
- Only large market size that have bimodal distribution
- Large market size can be subset into 2 groups based on Location ID. Location ID from 200 to 222 forms the most of the outliers and make better sales than the 900s group.
- I decide not to delete this outliers.

GSHEET LINK

2. Exploratory Data Analysis

Sales Performance Each Week





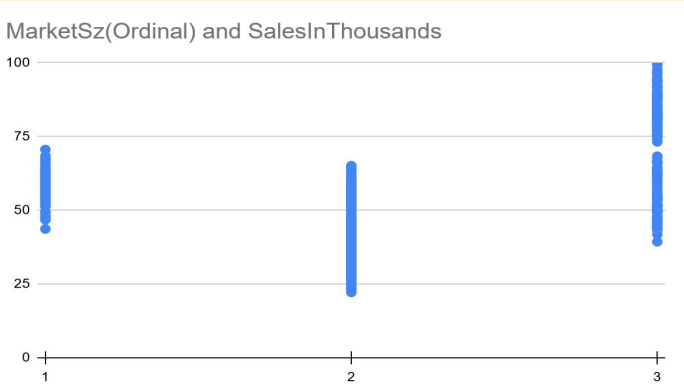
INSIGHTS:

- Campaign 1 sales outperform other campaign in every week.
- Large market size(loc ID 200+) sales is far better than other in every week

GSHEET LINK

3. Correlation Study

	MarketID	*MarketSz(Ordinal)*	*LocationID*	*AgeOfStore*	*Promotion*	*Week*	*SalesInThousands*
MarketID	*1.00*						
MarketSz(Ordinal)	0.26	1.00					
LocationID	1.00	0.27	1.00				
AgeOfStore	-0.05	-0.17	-0.05	1.00			
Promotion	-0.05	-0.06	-0.05	0.06	1.00		
Week	0.00	0.00	0.00	0.00	0.00	1.00	
SalesInThousands	-0.19	0.45	-0.19	-0.03	-0.06	-0.01	1.00



t-critical	11.86091683
p-value	0

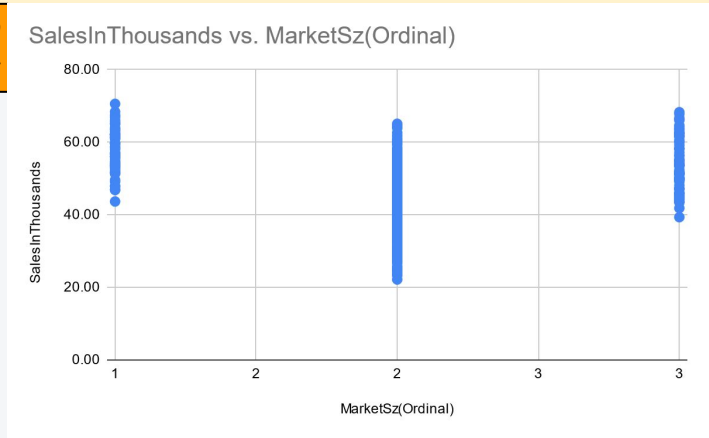
INSIGHTS:

- H0: there is no linear correlation between market size and sales (ρ=0)
- H1: there is linear correlation between market size and sales (ρ≠0)
- Conclusion: reject H0, as market size get bigger, sales also increased, with moderate positive correlation. However, this conclusion is generated while there is still Location ID 200s in the dataset

GSHEET LINK

3. Correlation Study

	MarketID	*MarketSz(Ordinal)*	*LocationID*	*AgeOfStore*	*Promotion*	*Week*	*SalesInThousands*
MarketID	1.00						
MarketSz(Ordinal)	0.67	1.00					
LocationID	1.00	0.67	1.00				
AgeOfStore	-0.10	-0.16	-0.10	1.00			
Promotion	-0.04	-0.10	-0.04	0.07	1.00		
Week	0.00	0.00	0.00	0.00	0.00	1.00	
SalesInThousands	0.30	-0.04	0.30	0.10	-0.14	-0.01	1.00



INSIGHTS:

- If I exclude Location ID 200s from the analysis, there is almost no linear correlation between market size and sales.

GSHEET LINK

3. T-Test on Sales Per Campaign

	Promo 1	*Promo 2*	*alpha*
Mean	51.96465278	42.976875	0.01667*
Observations	144	160	
P(T<=t) two-tail	0		
	Promo 1	*Promo 3*	
Mean	51.96465278	48.36407895	
Observations	144	152	
P(T<=t) two-tail	0.0009		
	Promo 2	*Promo 3*	
Mean	42.976875	48.36407895	
Observations	160	152	
P(T<=t) two-tail	0.0000002		

*alpha=0.01667 because of Bonferroni correction

INSIGHTS:

- Marketing campaign 1 better than other 2 campaign. 2nd best is campaign 3. All of the differences is statistically significant

GSHEET LINK

3. T-Test on Market Size Effect on Sales

	Small	*Medium*	*alpha*
Mean	57.40933333	43.98534375	0.0083*
Observations	60	320	
P(T<=t) two-tail	0		

	Medium	*Large(Loc900)*	*alpha*
Mean	43.98534375	53.77625	0.0083*
Observations	320	80	
P(T<=t) two-tail	0		

	Small	*Large(Loc900)*
Mean	57.40933333	53.77625
Observations	60	80
P(T<=t) two-tail	0.0023	

	Medium	*Large(Loc200)*
Mean	43.98534375	84.97170455
Observations	320	88
P(T<=t) two-tail	0	

	Small	*Large(Loc200)*
Mean	57.40933333	84.97170455
Observations	60	88
P(T<=t) two-tail	0	

	Large(Loc900)	*Large(Loc200)*
Mean	53.77625	84.97170455
Observations	80	88
P(T<=t) two-tail	0	

*alpha=0.0083 because of Bonferroni correction

INSIGHTS:

- Market size Large Location ID 200+ has the best sales with huge difference, followed by Small, Large Location ID 900+ and Medium. All the differences is statistically significant.

GSHEET LINK

4. Final Conclusion and Recommendation

- **Marketing campaign 1 yields the best sales**, however, the difference is small with campaign 3. Need to discuss what defines sizable effects and need to compare with cost to perform each marketing campaign.
- Market Size has different impact on Sales, **the rank is Large Location ID 200 > Small > Large Location ID 900 > Medium**.
- Company can **explore other characteristic** that determine the sales, especially on location ID 200.
- If the sampling method used by the company is cluster sampling, **it is better to do simple random sampling or stratified sampling** based on market size next time.

Thank You!