Boston University

AD 654 - Marketing Analytics

Professor Gregory Page

May 07, 2023

Case Study Report

Uttej Simha

Matthew Garber

Jing-Chian Teng

Shrutkirti Wagh

Juthathip Thasai

Sapna Shetty

Lobster Land Annual Summer Season

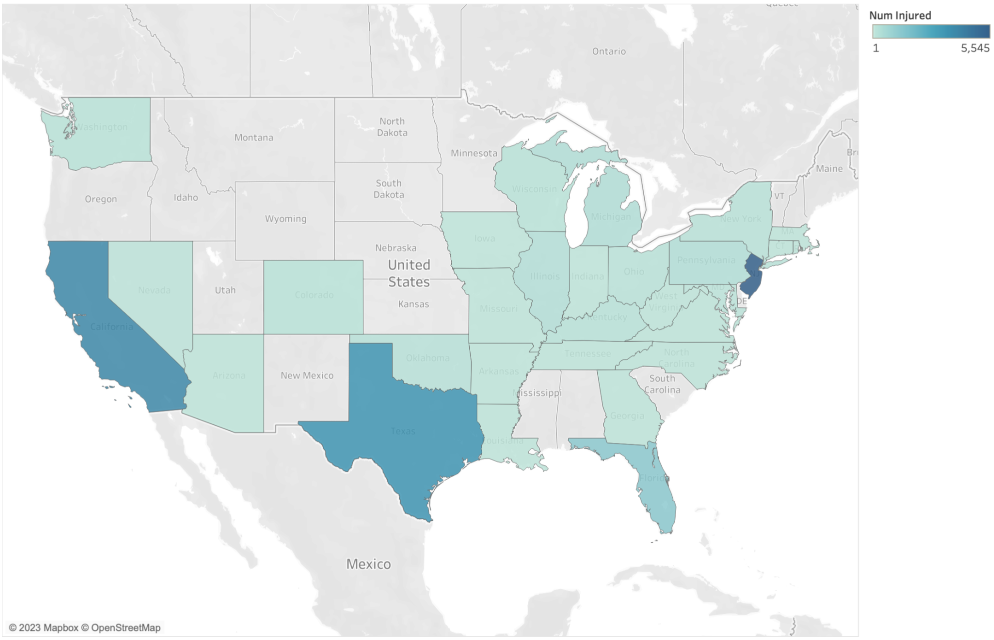
Data Visualization

Looking at the park accidents dataset, there are some good insights that tell a story about when and where these accidents take place. When looking at a visualization of accidents each year, there were very few prior to 1998, and with a steep uptick in the early 2000s, which has since tapered down leading into 2010. One could look at this and think amusement parks were incredibly safe in the nineties, and then the early 2000s had a popular movement of deregulating these parks, and then they finally got safe again. The reality of the situation is more likely that amusement parks grew in popularity, and thus there were more accidents with a higher volume of customers. We also speculate the standard for what qualifies as an accident to be reported may have gotten looser in the 2000s. For example in the 1990s, someone may have needed to have a serious injury for it to count, but in the early 2000s someone who stubbed their toe could make it into the dataset. We speculate that these parks saw this trend and gradually have made their rides safer, and have made air-tight waivers that protect themselves from any potential injury lawsuits.

By analyzing the injury data by state, we can identify states with higher rates of injury among our customer base. From the injured people from each state, we can notice those states have a higher number of injured, but we speculate that this is largely due to those states having more theme parks, and therefore more customers to get into an accident. It is unlikely that theme parks in California, New Jersey, and Texas are any more unsafe than Lobster Land

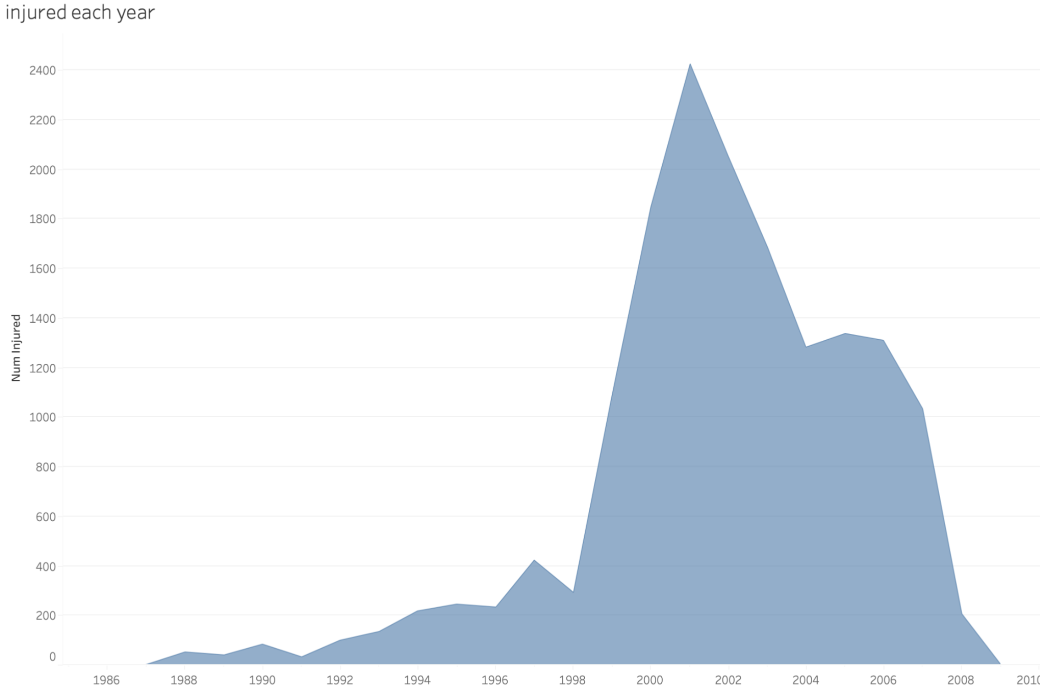
In the device category and injured section, we can understand the potential risk of certain attractions and rides. This is especially useful to Lobster Land because they can have a good idea of the risk profile when adding onto the park. Also, in high-risk attractions, there should be extra care and safety measures where necessary. In the Operational, Employee, and Mechanical error part, we can see most of the injuries are not related to employee, operation, and mechanical. However, there are some injuries related to the operation, employee, and mechanical, so if we want to eliminate those factors to decrease the number of injuries, we need to spend more time teaching our employees the procedures and checking the equipment more frequently.

injured from states



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Num Injured. Details are shown for Acc State.

injured each year



The plot of sum of Num Injured for Acc Date Year.

Device Category and injured

	Industry Sector			
Device Category	amusement ride	water park	recreation	unknown
coaster	3,111			
spinning	2,324			
water ride	1,253			
cars & track rides	1,062			
other attraction	453			
pendulum	431			
vertical drop	232			
water slide		3,653		
aquatic play		467		
wave device		189		
float attraction		187		
go-kart			1,798	
play equipment			404	
inflatable			264	
challenge activity			96	
trampoline			33	
alpine activity			45	
laser tag			1	
unknown				91

Num Injured: 1 – 3,653

Sum of Num Injured broken down by Industry Sector vs. Device Category. Color shows sum of Num Injured. The marks are labeled by sum of Num Injured.

Operational, Employee, Mechanical error and injury

		Mechanical	
Employee	Op Error	False	True
False	False	14,031	1,658
	True	328	34
True	False	10	3
	True	30	

Num Injured: 3 – 14,031

Sum of Num Injured broken down by Mechanical vs. Employee and Op Error. Color shows sum of Num Injured. The marks are labeled by sum of Num Injured.

Segmentation and Targeting

We assessed the Ski Hotels dataset with a clustering model in order to get insights on the different types of customers and what they want in their experience. These clusters were identified using features such as price, distance from lift, altitude, and total piste length. We can then take this segmentation into consideration when offering different deals that cater to customers and their varying spending habits and needs. We felt that a K-value of four, breaking up customers into four distinct groups was appropriate for the given data.

The Name of each clusters

Cluster 0 - Budget-Friendly Skiers:

This cluster has the lowest prices and low altitude, indicating that it is a more budget-friendly option for skiers. Resorts in this cluster likely have smaller ski areas with fewer runs and lifts, making them ideal for beginners and families on a budget. Offering more affordable equipment rentals, lessons, and packages will attract these budget-conscious skiers.

Cluster 1 - Family-oriented skiers:

This cluster has a moderate distance from the lift as well as moderate prices and altitudes, indicating a middle-tier option for skiers. Resorts in this cluster might have a mix of beginner and intermediate runs and lifts, making them ideal for families and groups with varying skill levels. Offering a wide range of equipment rentals and lessons will cater to the different needs of this generalist group.

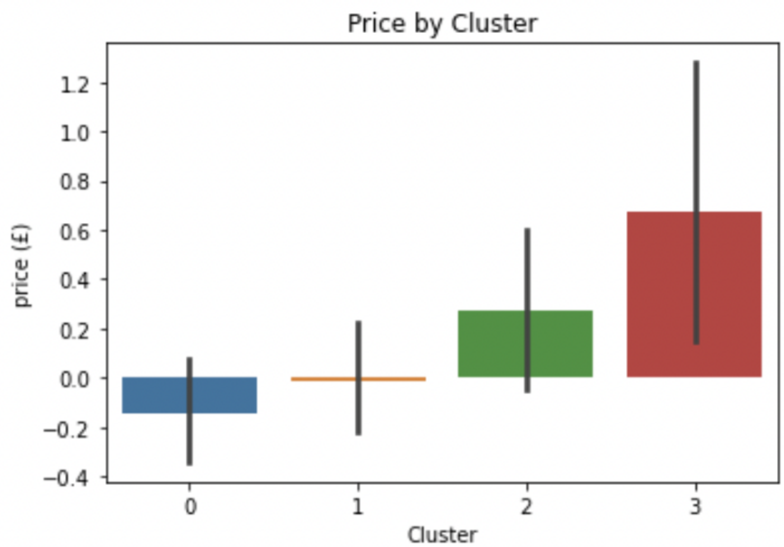
Cluster 2 - Advanced skiers:

This cluster has a larger proportion of difficult runs, indicating that it is a more advanced option for skiers. Resorts in this cluster might have a greater focus on challenging terrain and attract more experienced skiers. Offering high-performance equipment rentals and advanced lessons, if any lessons at all, to cater to the needs of these skilled skiers.

Cluster 3 - Luxury Skiers:

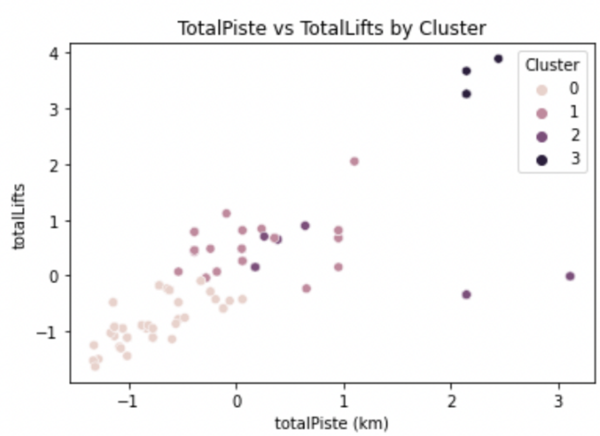
This cluster has the highest prices and a large number of lifts, indicating that these are the premium options for skiers. Resorts in this cluster might offer a wide variety of runs, including difficult terrain and multiple ski areas as well as the longest total runs. Offering luxury equipment rentals, luxury lodging, and personalized lessons to cater to the high-end needs of this high-spending group.

```python
sns.barplot(x="Cluster", y="price (£)", data=kmeans2)
plt.title("Price by Cluster")
plt.show()
```



This bar plot shows the average price (£) for each cluster. It reveals that Cluster 3 has the highest average price and tops the range, while Cluster 0 has the lowest average price and contains the bottom of the range.

```python
sns.scatterplot(x="totalPiste (km)", y="totalLifts", hue="Cluster", data=kmeans2)
plt.title("TotalPiste vs TotalLifts by Cluster")
plt.show()
```

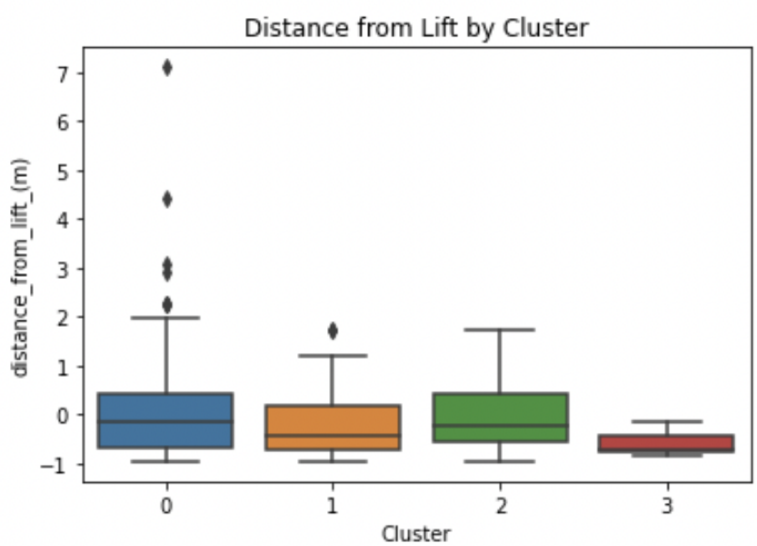


This scatterplot shows the relationship between the total piste length and total number of lifts for each cluster. It reveals Cluster 0 has the lowest values indicating lowest number of lifts and shortest post when compared to others, while cluster 3 is on the other end of the spectrum.

```python
[ ] g = sns.FacetGrid(kmeans2, col='Cluster', hue='Cluster', col_wrap=4)
    g.map(sns.histplot, 'totalRuns', kde=False)
    g.fig.subplots_adjust(top=0.9)
    g.fig.suptitle('Histogram of totalRuns')
    g.add_legend()
    plt.show()
```

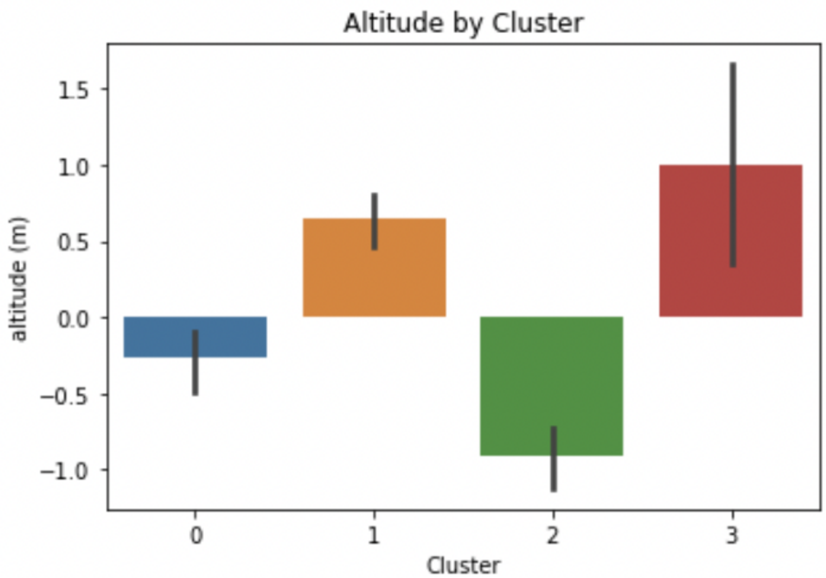
This histogram shows the distribution of totalRuns for each cluster. We can see that cluster 0 has a lower concentration of total runs, while cluster 3 has a higher concentration compared to the rest of the distribution.

```python
sns.boxplot(x="Cluster", y="distance_from_lift_(m)", data=kmeans2)
plt.title("Distance from Lift by Cluster")
plt.show()
```



The box plot shows the distribution of the "distance_from_lift_(m)" feature across the five clusters. We can see that cluster 0 has a higher median value for Distance from lift compared to the other clusters, as well as some resorts with very high values.

```
sns.barplot(x="Cluster", y="altitude (m)", data=kmeans2)
plt.title("Altitude by Cluster")
plt.show()
```



This bar plot shows the average altitude for each cluster. It reveals that Cluster 3 has the highest altitude, while Cluster 2 has the lowest altitude.

In conclusion, the segmentation analysis provides a valuable framework for ski resort vendors to understand their target customer segments and tailor their marketing strategies and offerings accordingly. By recognizing the distinct characteristics and preferences of each cluster, ski resorts can better meet the needs and expectations of their customers, leading to improved customer satisfaction and business success. Lobster Land can perform a similar analysis in order to know their own customer base better as well.

Forecasting Total Spending

	A	B	C	D	E
1		2019	2020	2021	2022
2	Total Revenue	9452	4307	5788	8773
3	COGS	7017	3724	4133	6075
4	Gross Profit	2435	583	1655	2698
5	Opex	859	702	638	604
6	Operating Income	1576	-119	1017	2094
7	Interest income exp	-414	-429	-397	-415
8	Other income expense	82	-376	-60	55
9	Pretax income	1244	-924	560	1734
10	Tax Provision	358	-204	153	477
11	Net Income	881	-715	410	1255

Hilton (HLT) csv.

	A	B	C	D	E
1		2019	2020	2021	2022
2	Total Revenue	5020	2066	3028	5891
3	COGS	4077	2067	2603	4603
4	Gross Profit	943	-1	425	1288
5	Opex	746	631	676	890
6	Operating Incom	197	-632	-251	398
7	Interest income (	-75	-128	-163	-150
8	Other income ex	884	-200	458	115
9	Pretax income	1006	-960	44	363
10	Tax Provision	240	-257	266	-92
11	Net Income	766	-703	-222	455

Hyatt (H) csv.

```python
hlt_ni = hlt.iloc[[9]]
hlt_ni
```

	Unnamed: 0	2019	2020	2021	2022
9	Net Income	881	-715	410	1255

```python
#Find Compound Annual Growth Rate
cagr = (1255-881)**(1/4)-1
cagr
```

```
3.3976220398999404
```

```python
projection = pd.DataFrame(columns=['Year', 'Net Income'])

ni = 1255

for i in range(1,4):
  proj_year = int(2022 + i)
  projected_ni = round(ni * (1+ (cagr/100))**i)
  projection.loc[i-1] = [proj_year, projected_ni]
```

```python
projection
```

	Year	Net Income
0	2023	1298
1	2024	1342
2	2025	1387

Hilton Projections

```python
h = pd.DataFrame(h)
h_ni = h.iloc[[9]]
h_ni
```

	Unnamed: 0	2019	2020	2021	2022
9	Net Income	766	-703	-222	455

```python
#Find Compound Annual Growth Rate (From 2021 to 2022)
cagr1 = (455+222)**(1/4)-1
cagr1
```

```
4.100904200484006
```

```python
projection1 = pd.DataFrame(columns=['Year', 'Net Income'])

ni1 = 455

for i in range(1,4):
  proj_year1 = int(2022 + i)
  projected_ni1 = round(ni1 * (1+ (cagr1/100))**i)
  projection1.loc[i-1] = [proj_year1, projected_ni1]

projection1
```

	Year	Net Income
0	2023	474
1	2024	493
2	2025	513

Hyatt Projections

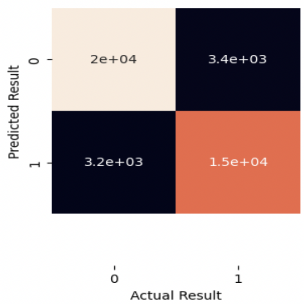
For the forecasting analysis, we decided to keep it relatively simple, as these numbers can be very challenging to correctly pick, with many levers impacting that bottom line of net income. All of the financial information was taken from Yahoo Finance, and moved to a CSV file. From there, the file was read in, and we created a new dataframe to isolate net income.

The method we used was calculating an appropriate compound annual growth rate of net income from 2019-2022, and then applying that three years out into the future to get net income for 2023, 2024, and 2025.

One important factor to note here is the impact of the Covid-19 pandemic on the earnings of both Hyatt and Hilton. Being hotels, there was a big hit to revenue in 2020 and 2021, which then turned into negative or low earnings due to both of their high-fixed costs of physical locations. With that being said, earnings were relatively normal in 2022, so we felt that for Hilton the CAGR for all four years worked. For Hyatt, however, their 2022 earnings were lower than 2019, but had a positive trend from 2021, so we decided to only use net income from 2021 to 2022. This year over year growth was very abrupt, so we decided to halve that growth rate to have more realistic projections for the next three years. The final numbers are expressed in millions of dollars.

Classification

```
# confusion martix
mat = confusion_matrix(pred, y_test) #the first value given here will show up on the y-axis
sns.heatmap (mat, square=True, annot=True, cbar=False)
plt.xlabel("Actual Result")
plt.ylabel("Predicted Result")
a, b = plt.ylim()
a += 0.5
b -= 0.5
plt.ylim(a, b)
plt.show()
```




```
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, balanced_accuracy_score

rf_model.fit(X_train, y_train)
# Make predictions on the testing data
y_pred = rf_model.predict(X_test)

# Calculate and print confusion matrix
cm = confusion_matrix(y_test, y_pred)
print("Confusion matrix:")
print(cm)

# Calculate and print accuracy rate
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy rate:", accuracy.round(4))

# Calculate and print sensitivity rate
sensitivity = recall_score(y_test, y_pred)
print("Sensitivity rate:", sensitivity.round(4))

# Calculate and print specificity rate
specificity = cm[0,0]/(cm[0,0]+cm[0,1])
print("Specificity rate:", specificity.round(4))

# Calculate and print precision
precision = precision_score(y_test, y_pred)
print("Precision:", precision.round(4))

# Calculate and print balanced accuracy
balanced_accuracy = balanced_accuracy_score(y_test, y_pred)
print("Balanced accuracy:", balanced_accuracy.round(4))
```

```
Confusion matrix:
[[21496  1942]
 [ 2464 15660]]
Accuracy rate: 0.894
Sensitivity rate: 0.864
Specificity rate: 0.9171
Precision: 0.8897
Balanced accuracy: 0.8906
```

Using a logistic regression model, we can learn which factors significantly impact customer satisfaction. Management can use this information to improve their services to enhance customer satisfaction, and to capture more revenue. The model's feature selection process revealed that the five most important factors influencing customer satisfaction are the type of booking, type of travel, hotel wifi service, Stay comfort,and common room entertainment. This information can help hotel managers to prioritize the services they offer based on their impact on customer satisfaction.

Next, we used a confusion matrix to derive performance metrics. Our model had an accuracy rate of 89.4%, which means that it can correctly predict customer satisfaction in almost 9 out of 10 cases. The sensitivity rate is 86.4%, indicating that it can correctly identify satisfied customers, while the specificity rate of 91.7% shows that it can accurately predict dissatisfied customers. The precision rate of 88.9% implies that our model can correctly identify satisfied customers among all the customers it predicts as satisfied.

Overall, the results of this logistic regression model can help hotel managers to better understand the factors that influence customer satisfaction and tailor their marketing strategies to attract and retain customers. They can use the information to improve their services in areas that have the most significant impact on customer satisfaction and prioritize the services they offer based on their importance. By doing so, they can enhance customer satisfaction and, in turn, increase customer loyalty and drive revenue higher.

Conjoint Analysis

The first conjoint analysis we performed was to see what amenities hotel guests would prefer. The above feature importance table shows the strength of each feature in predicting the outcome variable. The higher the importance value, the more influential the feature is in our model.

From the table, we can see that the most important features are "jacuzzi_Yes" and "jacuzzi_No", followed by "gym_Advanced", "gym_Super", "gym_None", and "gym_Basic". The importance values of these features are much higher than the rest of the features, indicating that they have a strong influence on the outcome variable.

On the other hand, the features "pool_temp_84", "pool_temp_80", and "pool_temp_76" have negative importance values, which means that they have a negative impact on the outcome variable.

Overall, the feature importance table can be used to identify the most important features and help us understand which features to focus on when improving the model's performance.

The next analysis we performed was to use a cost constraint of $250 per room per night, and our conjoint model results suggest that the most significant factor affecting the per-serving cost is the presence or absence of a shuttle bus. The presence of a shuttle bus increases the per-serving cost by $75.00, while the absence of a shuttle bus has no effect on the cost. Other factors that significantly affect the per-serving cost include parking, breakfast options, and gym options.

Based on these results and the cost constraint of $250 per room per night, the recommended set of amenities would include no shuttle bus, an open lot parking, continental breakfast, and a basic gym. These amenities would keep the per-serving cost at a reasonable level while still offering some desirable options for guests.

Summary Statistics

In this section, we explored our dataset in order to find some insights about park injuries. The dataset contains 14,884 observations. The acc_id variable is a unique identifier for each accident. The num_injured variable is the number of people injured in the accident. The age_youngest variable is the age of the youngest person injured in the accident. The year variable is the year in which the accident occurred. On average, there were 1.08 injured individuals per accident. The youngest person injured in an accident was, on average, 16.86 years old. The accidents in this dataset occurred between 1986 and 2009.

simple groupby()

```
[ ] # Count accidents by state
    df_grouped_state = df.groupby('acc_state').size()
    # Sum injured people by year
    df_grouped_year = df.groupby('year')['num_injured'].sum()
```

complex groupby()

```
[ ] # Complex group by functions
    df_grouped_device_type = df.groupby(['device_type', 'gender']).agg({'num_injured': 'sum', 'acc_id': 'count'})
    # Group by device type and gender, sum the number of injured people and count the number of accidents
    df_grouped_city_state = df.groupby(['acc_state', 'acc_city']).agg({'num_injured': ['sum', 'mean', 'std'], 'age_youngest': 'median'})
    # Group by state and city, calculate the sum, mean, and standard deviation of injured people and median of youngest age
    df_grouped_bus_sector = df.groupby(['bus_type', 'industry_sector']).agg({'num_injured': ['sum', 'mean'], 'acc_id': 'count', 'year': ['min', 'max']})
    # Group by business type and industry sector, sum the number of injured people and count the number of accidents, and calculate the minimum and maximum year of the accidents
```

To look at more targeted areas of the data, we used group by to isolate what we wanted. While there are more examples in the code, a couple of our insights were the following. Counting accidents by state: The output shows the number of accidents that occurred in each state, with the number of accidents ranging from 1 to 5,646. California (CA) had the highest number of accidents (3,407), followed by New Jersey (NJ) with 5,646 accidents.

Summing injured people by year: This output shows the total number of injured people in accidents for each year, ranging from 1 to 2,424. The number of injuries appears to have been highest in the year 2002, with 2,051 injuries, while the lowest number of injuries occurred in the year 2009, with only two injuries.

A/B Statistical Testing

We performed an A/B test to decide which picture Lobster Land should use in their emails. Based on the results we obtained from the ANOVA analysis, it can be concluded that the picture type has a significant effect on the registration rates. The cost per conversion and the demographic profile of the target audience were also computed for each picture type. Among the different pictures, the sunset one appears to have the best performance. Therefore, it is recommended to use the sunset picture for future promotional campaigns to maximize registration rates and optimize cost per conversion. However, it is important to note that other factors, such as the timing and placement of the ad, should also be considered when designing promotional campaigns.

	recipient	site_duration
pic_seen		
Main St	1148	24.294059
Sunset	1110	22.949189
Waterslide	1142	10.975394

Strategic Memo

Introduction:

The Golden Arch Hotel was a unique venture by McDonald's Corporation that opened in Rumlang, Switzerland, in 2001. The hotel was located near the Zurich airport and was marketed towards business travelers and families. The hotel's design incorporated McDonald's branding, with yellow and red decor and a restaurant that served McDonald's food. The Golden Arch Hotel was an attempt by McDonald's to diversify its business and expand into the hospitality industry. The purpose of this case study is to analyze the strengths, weaknesses, opportunities, and threats (SWOT) of McDonald's as it entered the hotel industry, evaluate its branding and positioning strategy, and provide recommendations for Lobster Land to expand its business.

SWOT Analysis for Golden Arches:

Strengths:

McDonald's had several strengths when entering the hotel industry, including its established brand recognition and reputation worldwide. McDonald's had strong financial resources for investment, proven ability to operate successful franchises, and a strong focus on customer service. McDonald's also had an experienced management team with expertise in the foodservice industry.

Weaknesses:

However, the Golden Arches also had several weaknesses that could hinder their success in the hotel industry. McDonald's had limited experience in the hotel industry, and this lack of experience could translate into lower quality customer service or operational challenges. Additionally, McDonald's had limited flexibility in branding and menu offerings, which could limit the hotel's appeal to a wider range of guests. The company's dependence on the fast-food market, which may not translate to the hotel industry, is another potential weakness.

Opportunities:

There are several opportunities that the Golden Arches could leverage when entering the hotel industry. Growing demand for affordable hotel accommodations presents a significant opportunity for the company. Additionally, the potential for cross-selling opportunities with McDonald's fast-food chain could allow the company to tap into a wider range of customers. Finally, the potential to leverage McDonald's existing supply chain for hotel operations could provide significant cost savings for the company.

Threats:

However, there are also several threats that the Golden Arches must consider when entering the hotel industry. High competition in the hotel industry from established players could make it difficult for McDonald's to gain market share. Dependence on the success of individual hotel locations is another potential threat. Finally, there is the potential for cannibalization of sales from McDonald's fast-food chain, as the hotel's restaurant served McDonald's food.

Branding and Positioning:

From a branding and positioning perspective, McDonald's did well by leveraging its well-known brand name and colors, using a similar theme and decor in the hotel. By doing so, McDonald's was able to attract customers who were already familiar with the brand and its fast-food chain. However, McDonald's could have done better by differentiating its hotel business from its fast-food chain. The hotel's decor, amenities, and menu could have been more upscale to attract a higher-end clientele. The hotel could also have focused on eco-friendliness or incorporated local cultural elements to appeal to a wider range of guests.

Recommendations for Lobster Land:

Based on the case study, Lobster Land is looking to expand its business. Here are some recommendations that could help the company succeed in its expansion efforts:

- Conduct market research to identify areas with a high demand for seafood and assess the competition in those markets. This will help Lobster Land to identify the best locations for expansion and ensure that there is demand for their seafood offerings.
- Develop a strong brand identity and differentiate the restaurant's offerings from competitors by creating a unique ambiance and menu.
- Offer a variety of menu options that cater to different dietary preferences and price points. Lobster Land could consider offering options or creating a separate menu for budget-conscious customers.
- Leverage technology to improve the customer experience and increase efficiency. Lobster Land could consider implementing online ordering and payment systems or investing in kitchen automation technology to reduce wait times and increase order accuracy.
- Invest in employee training and development to ensure consistent quality of service and food offerings. Lobster Land could create a comprehensive training program for new employees or offer ongoing professional development opportunities to improve retention and employee satisfaction.

  Example: In Golden Arch you see a review of one of the customers saying there was no one present at the hotel reception. This professional consciousness training should be a key priority here.
- Finally, Lobster Land could consider strategic partnerships or collaborations with other businesses to expand its reach and customer base. For example, partnering with local hotels or tourist attractions could increase exposure and attract new and local customers.

Conclusion for Golden Arches:

In conclusion, the Golden Arch Hotel by McDonald's Corporation was an interesting attempt to diversify its business and enter the hospitality industry. McDonald's had several strengths, weaknesses, opportunities, and threats to consider when entering this new market. From a branding and positioning perspective, McDonald's leveraged its well-known brand name and colors, but could have done better by differentiating its hotel business from its fast-food chain. Finally, Lobster Land can learn from this case study and leverage market research, technology, and strategic partnerships to expand its business and differentiate itself from its competitors.

Conclusion

After performing our analysis, we drew many high-quality insights into park accidents as well as hospitality company earnings and potential strategies to make a horizontal move to enter the hotel business. We also learned a lot about types of hospitality industry customers, what sets of features to offer those customers, what draws them to promotions, and predicting their satisfaction.

We were able to make these insights by exploring the data visually and statistically, performing both cluster and conjoint analysis, utilizing A/B tests, and classification. We are confident that Lobster Land can make use of our findings whether they want to make strategic changes, or just want to get to know their customers better.