Data-Driven Insights for Disney's Strategic Decisions

A Comprehensive Marketing Analytics Study

Tara Cool - Spring 2023

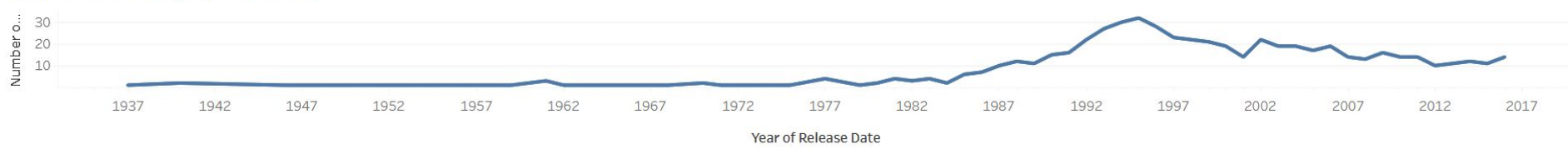# Objectives

1. **Data Visualization (Tableau)**
2. **Summary Stats**
3. **Segmentation & Targeting**
   a. **k-Means Clustering Algorithm**
4. **Conjoint Analysis**
   a. **Multiple Linear Regression Model**
5. **Strategic Analysis Memo**
6. **Forecast - Net Income**
   a. **Simple Moving Average**
7. **Classification**
   a. **Logistic Regression Model**
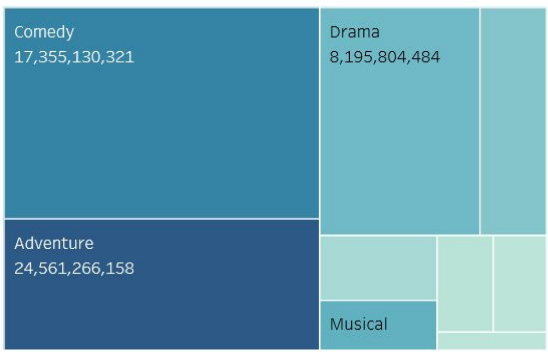   b. **Random Forest Algorithm**
8. **A/B Testing**

Exploratory Data Analysis: Data Visualization + Summary Stats

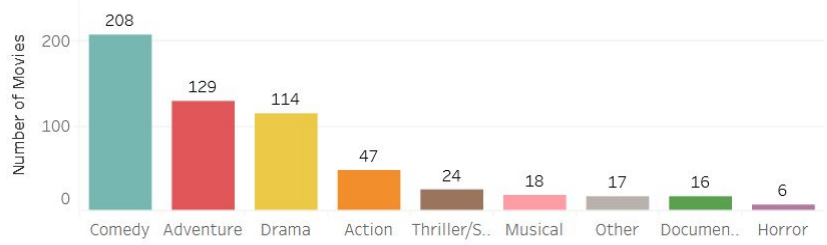Exploring Disney's Film Production Segment
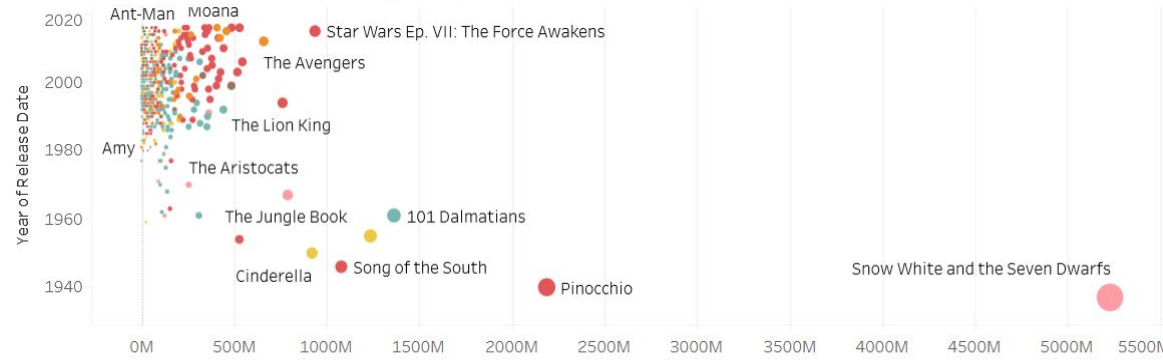
Time-Series of Movie Releases
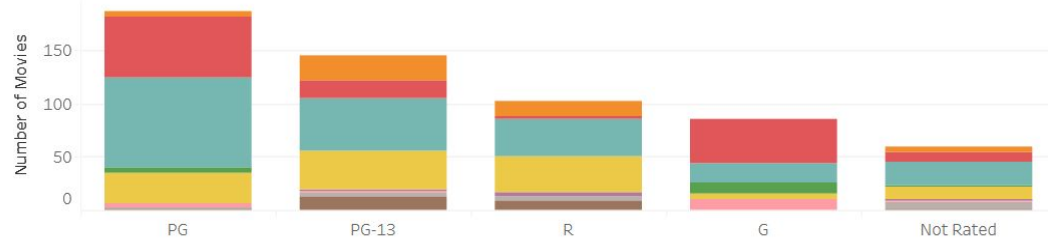

Total Revenue by Genre

Comedy 17,355,130,321	Drama 8,195,804,484	
Adventure 24,561,266,158		
	Musical	

Inflation A... 140,483,092 ▬▬▬▬ 24,561,266,158

Release Date vs. Inflation-Adjusted Gross Revenue

Ant-Man Moana
● Star Wars Ep. VII: The Force Awakens
The Avengers
The Lion King
Amy
The Aristocats
The Jungle Book   ● 101 Dalmatians
Cinderella   ● Song of the South   Snow White and the Seven Dwarfs
● Pinocchio

Genre ■ Action ■ Adventure ■ Comedy ■ Documentary ■ Drama ■ Horror ■ Musical ■ Other ■ Thriller/Suspen...

Distribution of Movies by Genre

| Comedy 208 | Adventure 129 | Drama 114 | Action 47 | Thriller/S... 24 | Musical 18 | Other 17 | Documen... 16 | Horror 6 |

Count of Movies by MPAA Ratings & Genre

PG   PG-13   R   G   Not Rated

Summary Statistics - Average *Revenue by Genre & MPAA Rating*

Historically & on average...
- *Musicals* generate more revenue than any other genre.
- *G-rated* films (i.e., all ages welcome) tend to generate the most revenue.
- *Documentaries* & *R-rated* films tend to underperform.

mpaa_rating	inflation_adjusted_gross
G	2.912610e+08
PG-13	1.029486e+08
PG	1.015414e+08
Not Rated	7.047577e+07
R	5.530581e+07

genre	inflation_adjusted_gross
Musical	6.035979e+08
Adventure	1.903974e+08
Action	1.374734e+08
Thriller/Suspense	8.965379e+07
Comedy	8.466773e+07
Romantic Comedy	7.777708e+07
Western	7.381571e+07
Drama	7.189302e+07
Concert/Performance	5.741084e+07
Black Comedy	5.224349e+07
Horror	2.341385e+07
Other	2.162373e+07
Documentary	1.271803e+07

Summary Statistics - *Distribution of Movies by Genre & MPAA Rating*

Most common genre + MPAA rating combination :

1. Comedy + PG
2. Adventure + PG

mpaa_rating genre	G	Not Rated	PG	PG-13	R	Total
Action	0	4	5	19	12	40
Adventure	42	10	57	17	3	129
Black Comedy	0	0	0	0	3	3
Comedy	17	21	77	37	30	182
Concert/Performance	2	0	0	0	0	2
Documentary	10	1	5	0	0	16
Drama	5	11	28	37	33	114
Horror	0	1	0	1	4	6
Musical	9	2	4	1	0	16
Other	0	7	2	4	4	17
Romantic Comedy	1	1	7	12	2	23
Thriller/Suspense	0	1	1	13	9	24
Western	0	0	1	4	2	7
Total	86	59	187	145	102	579

Significant potential for growth, given low production volume & high revenues (*See previous slide*).

Segmentation & Targeting

Disney Theme Parks Visitors

Segmentation - Modeling Process + Output

<u>Goal:</u> Segment families based on various demographic & financial attributes to understand their behavior & needs better.

k-Means Clustering Model used to group families into 'k' clusters
- Elbow method leveraged to determine optimal 'k' value (i.e., k= 6)
- Each cluster exhibits distinct characteristics regarding income, net worth, family size, travel, & leisure spending, providing valuable insights for targeted marketing or service offerings.

Cluster	est_net_worth mean	est_net_worth median	est_inc mean	est_inc median	number_visits mean	number_visits median	number_children mean	number_children median	oldest_child_age mean	oldest_child_age median	online_merch_avg mean	online_merch_avg median	est_annual_travel mean	est_annual_travel median	est_annual_leisure mean	est_annual_leisure median	est_hhold_FICO mean	est_hhold_FICO median	count
0	588381.67	466414.0	487670.81	329380.0	2.11	2.0	2.23	2.0	12.03	12.0	1839.08	1802.0	7893.73	6868.0	7560.13	6194.0	750.11	750.0	3057
1	6020044.02	5721637.0	3568802.52	3536832.5	2.73	1.0	2.16	2.0	12.19	12.0	1844.60	1806.0	19788.67	22603.0	43335.41	38666.5	747.38	748.0	372
2	56898.42	43243.0	42883.58	33232.5	2.14	2.0	1.82	2.0	12.80	13.0	799.28	783.0	2694.72	1917.0	2555.12	1891.0	763.52	764.0	2398
3	1821557.02	1821389.0	1289318.53	1278440.0	2.40	2.0	2.25	2.0	11.85	12.0	1840.54	1802.0	30435.94	28532.0	29966.14	27115.0	750.21	750.0	985
4	364745.48	136702.0	305458.61	100632.5	13.98	12.0	2.01	2.0	12.17	12.0	1325.25	1024.5	6168.07	4169.5	5863.48	3551.0	752.40	751.0	380
5	61325.55	45551.0	45954.21	35157.5	2.15	1.5	2.73	3.0	10.95	11.0	829.11	789.0	2856.81	2045.0	2688.86	2001.5	735.25	736.0	2058

Segmentation - Cluster Visualization

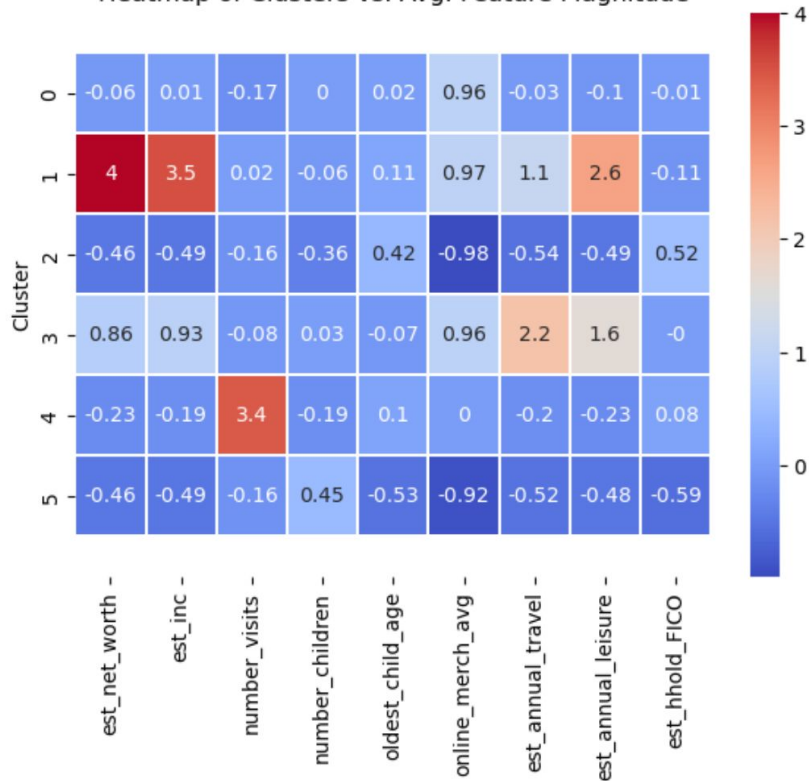Distribution of Oldest Childs Age by Cluster
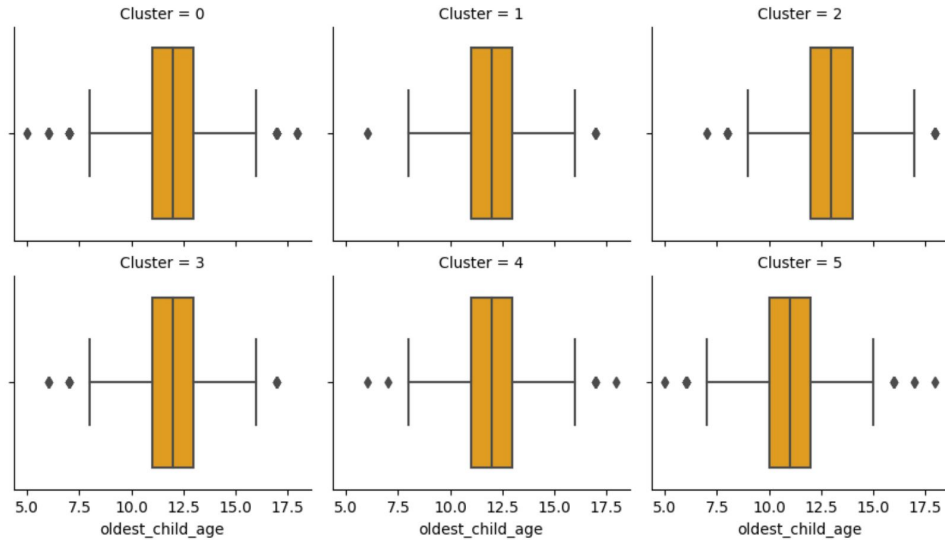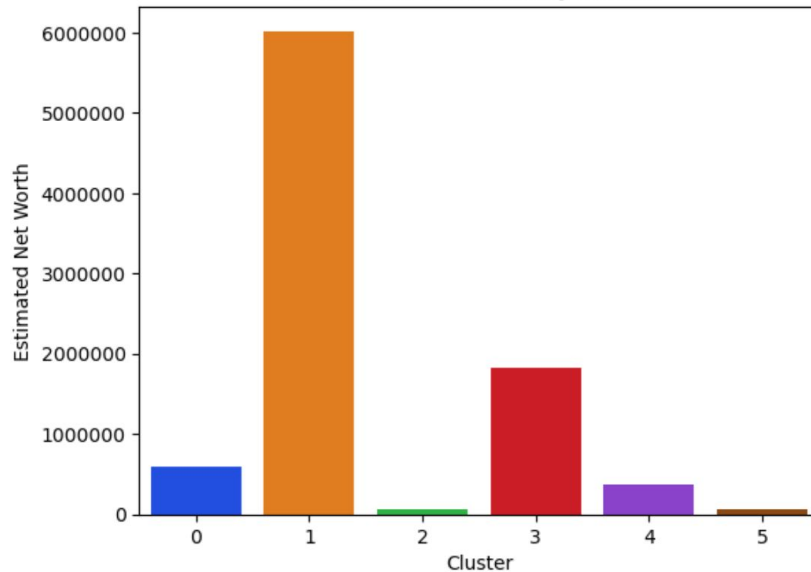
Distribution of Visit Count by Cluster
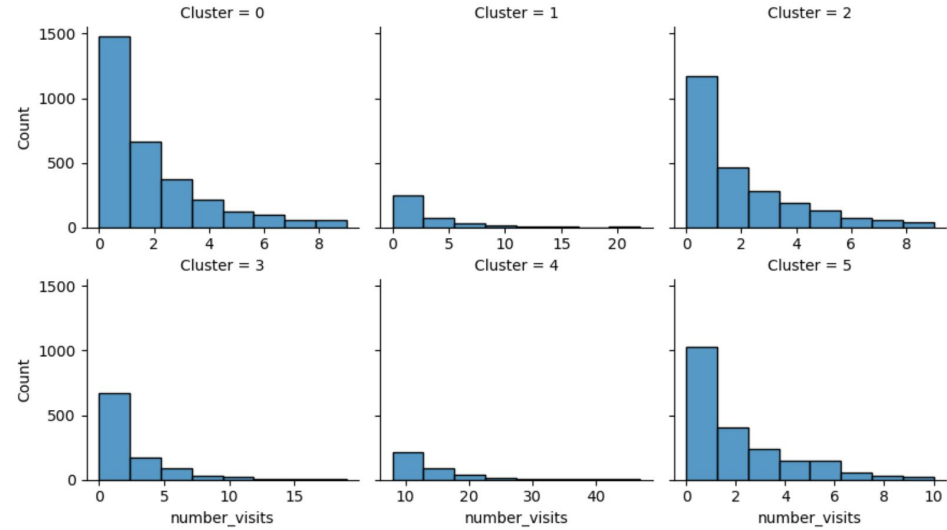
Segmentation - Cluster Visualization

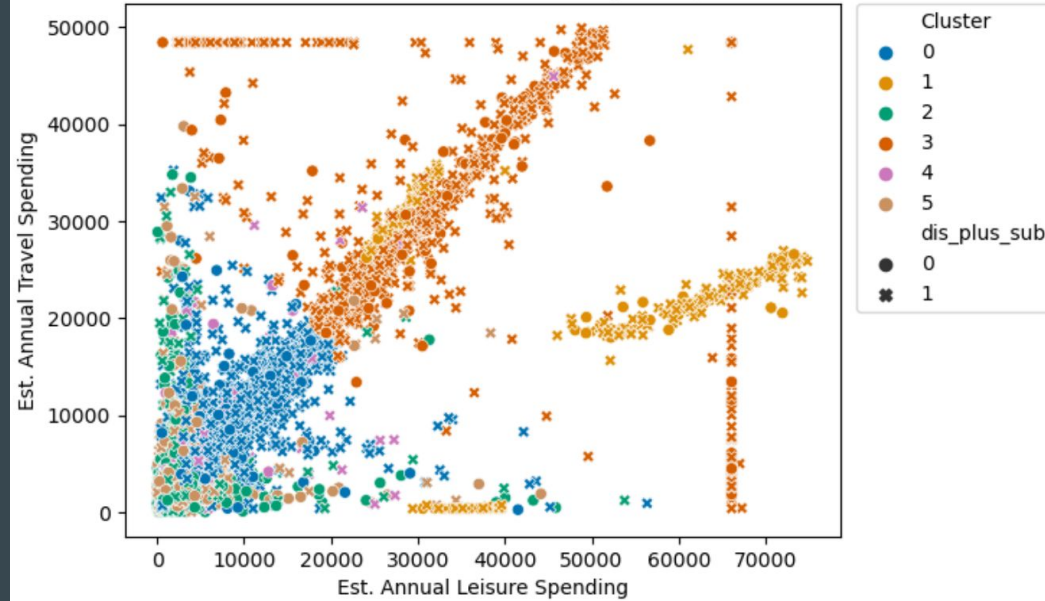Heatmap of Clusters vs. Avg. Feature Magnitude
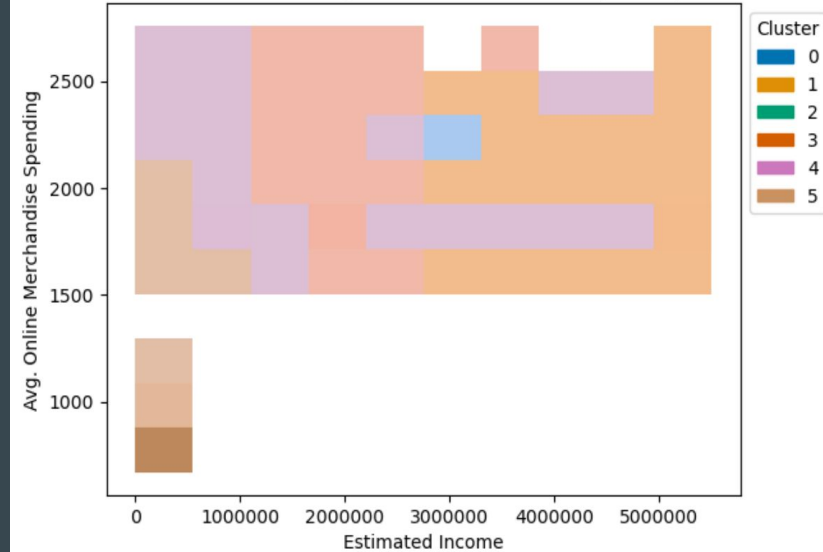
Estimated Net Worth by Cluster

Segmentation - Cluster Visualization

Naming Clusters + Segment Descriptions

★ Cluster 0: *"Lost Boys of Neverland"*
- Lowest avg. # of visits; the majority of these families have 'never' visited a Disney park/resort.

★ Cluster 1: *"Aristocats"*
- Highest avg. est. net worths, est. income, & avg. FICO (credit) score
- Spend more than any other group on online merchandise & annual leisure fees on average.

★ Cluster 2: *"Honey I Shrunk the Kids"*
- Lowest avg. est. net worth & est. income
- Spend the least on online merchandise, annual travel fees, & annual leisure fees on average
- Tend to have less children & the avg. oldest child in these families tends to be older than the oldest children of families in the other 5 clusters.

★ Cluster 3: *"Planes"*
- Greatest avg. est. annual travel spending
- Avg. est. net worth & est.income greater than most of the other clusters (i.e., besides cluster 1).

★ Cluster 4: *"Disney Dreamers"*
- Visit Disney with a greater frequency than the other clusters.

★ Cluster 5: *"Cheaper by the Dozen"*
- Lowest avg. est. household FICO (credit) score
- Lower avg. est. net worth & est. income than most of the other clusters (i.e., besides cluster 2).
- Greatest number of kids on avg. & the oldest child in these families tends to be much younger on avg. compared to other clusters

Targeting

Lost Boys of Neverland   Digital ads on streaming platform such as Disney+ or on the Disney website. Such an ad might highlight the most frequented attractions by first time/newer visitors.

Aristocats   Any luxurious attractions/resorts. For example, sending these customers email promotions for Disney's resort in Hawaii (e.g., maybe these promotions feature the resorts' pool, spa, or a relaxing game of golf).

Honey I Shrunk the Kids   Nostalgic rides/attractions (e.g., like the Honey I Shrunk the Kids that was previously in Disney World). Disney could host a special 12+ event for families with older children to explore the park/these new nostalgic attractions.

Planes   Disney-branded credit cards to earn points on as they spend. For example, for any travel expenditures, families can earn 2x the points, while in any Disney stores, parks, websites, or streaming platforms for 5x the points.

Disney Dreamers   Seasonal parking passes that may be used in any of the parks; everyday a family uses the parking pass, they receive 5-10% off ticket prices.

Cheaper by the Dozen   Coupons to the parks or promotional discounts (e.g., buy one, get one free kids tickets).

Conjoint Analysis

Optimal Amenities for Disney's New Hawaiian Hotel

Modeling Process

Goal: Identify the optimal set of amenities for Disney's to maximize customer satisfaction while staying within a $250 cost threshold per room, per night.

Data Preprocessing: All categorical variables were encoded as binary "dummy" variables for our analysis.

Multiple Linear Regression model was constructed with a baseline amenity package:
- ○ wifi=basic, breakfast=continental, parking=open_lot, gym=advanced, No added features (i.e., flex check, shuttle bus, air purifier, jacuzzi & VIP shop), pool temp=76
- Answers the question: "*How do modifications in amenities influence the average customer rating?*"

Model Output

Interpreting the Linear Regression Model:

The **intercept value** represents the model's *predicted customer rating* under the baseline amenity package & a total cost of $75.

```
2   lm = LinearRegression()
3   lm.fit(X_train, y_train)
4   lm.intercept_

5.4754473710162275
```

The model's coefficients represent the *change in customer ratings* (inc./dec.) when a specific amenity changes.

	Coefficient
WiFi_Network_Best in Class	1.724893
WiFi_Network_Strong	1.160978
breakfast_Full Buffet	0.543154
breakfast_None	-0.206894
parking_Valet	0.117816
gym_Basic	-0.155026
gym_None	-0.201128
gym_Super	0.030483
flex_check_Yes	0.466714
shuttle_bus_Yes	0.462803
air_pure_Yes	0.089652
jacuzzi_Yes	0.192366
VIP_shop_Yes	0.233033
pool_temp_80	0.090348
pool_temp_84	0.271793

Model Output CONT'D.

Cost-efficiency analysis:

- Model output combined with cost data.

- *Ratings change per dollar* computed for each amenity.

  a. Rating change coefficients / Total amenity cost

  b. Quantifies impact of $1 investment on ratings.

- Investigate data to identify optimal amenity bundle within $250 per room, per night cost constraints.

Rating Change/$ by Amenity
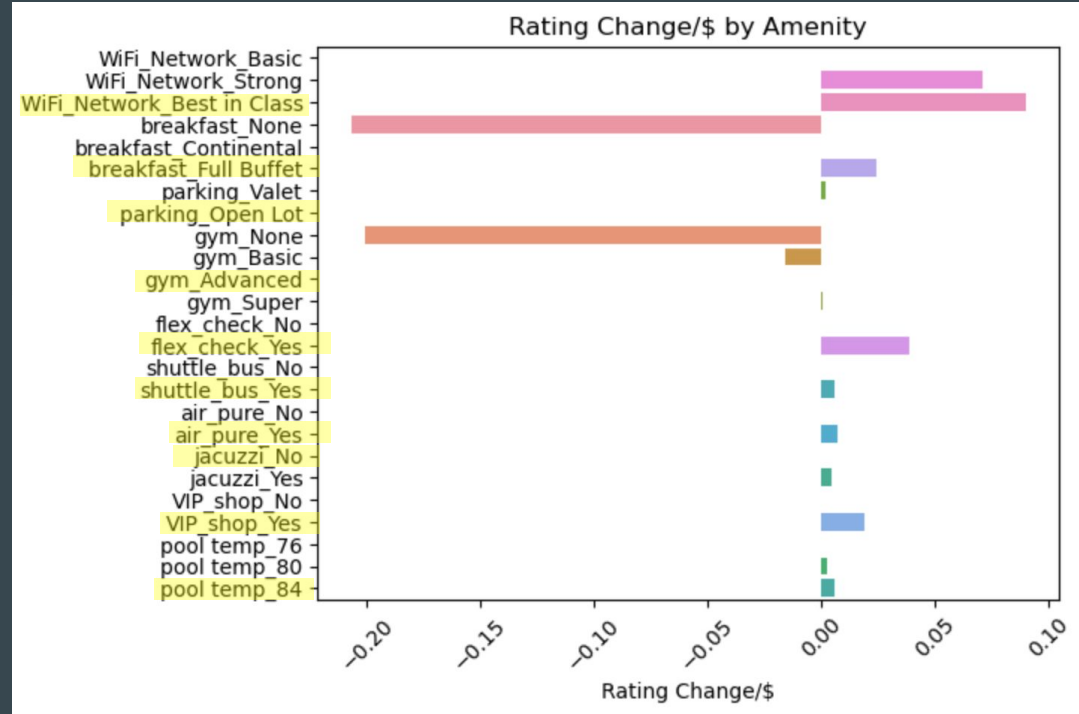
Recommendations

Balancing budget constraints and customer preferences, our **recommended amenities** for Disney's new Hawaiian hotel represent the *optimal set that maximizes satisfaction within a $250 cost threshold* per room, per night, ensuring a memorable and cost-effective experience for guests:

	Amenities	Cost	Rating addition
0	Wi-Fi = Best in Class	$19.15	1.724893
1	Breakfast = Full Buffet	$22.45	0.543154
2	Parking = Open Lot	$15	0
3	Gym = Advanced	$35	0
4	Flex Check = Yes	$12	0.089652
5	Shuttle Bus = Yes	$75	0.462803
6	Air Pure = Yes	$12.85	0.089652
7	Jacuzzi = No	$0	0
8	VIP Shop = Yes	$12	0.233033
9	Pool Temp = 84	$45	0.271793

Total Amenity Cost: $248.45 per room, per night
Projected Avg. Customer Rating: 8.89

Strategic Memo

Disney Theme Parks Division

Strategic Recommendations for Disney's Theme Parks Future

Case study: *"The Walt Disney Company: Theme Parks"*

<u>Purpose</u>: Provide Disney with actionable recommendations based on the analysis of the strategic decisions.

<u>Scope</u>: Closed universe analysis, drawing solely from the case study.

<u>Challenges:</u> Sustaining appeal to new vacationers while preserving Disney's essence.

<u>Decision Points</u>: (1) Invest in advanced technology through a park-wide upgrade, (2) introduce a standalone Hawaiian hotel, or (3) venture into the themed indoor recreation space with DisneyQuest?

Strategic Recommendations

Implement RFID-enabled Magic Bands:

1. Technological overhaul for immersive guest experiences
   a. Seamlessly integrate digital experiences into the physical realm of the parks.
2. Synergy & data integration
   a. Cohesive digital infrastructure for operations & data analysis.
3. Personalization & engagement
   a. Tailored guest experiences based on digital data (e.g., **engagement with attractions, transactions & access to personalized content**).
4. Operational efficiency & cost reduction
   a. Streamlining operations & reducing costs through digital transformation.
5. Maintain core Disney experience
   a. Ensuring technological enhancements complement Disney's core values.

Predictive Modeling

Forecasting Net Income

Modeling Process

Simple Moving Average (SMA)

<u>Goal:</u> Forecasting Disney's 2023 net income using historical net income data spanning from 2010 to 2022

Data Preparation:

- Remove unnecessary columns, rename columns, & convert data types.

Visualizing Historical Trend:

- Plotted historical net income trend from 2010 to 2022 to visualize the data distribution.
- Computed and plotted a 7-quarter rolling average to analyze long-term trends and smooth out fluctuations.

Forecasting for 2023:

- Created a DataFrame to store forecasts for the four quarters of 2023.
- Calculated the rolling 4-quarter average as a basis for forecasting net income for 2023.

Model Output

Net Income - Historical vs 4-Quarter Rolling Avg.
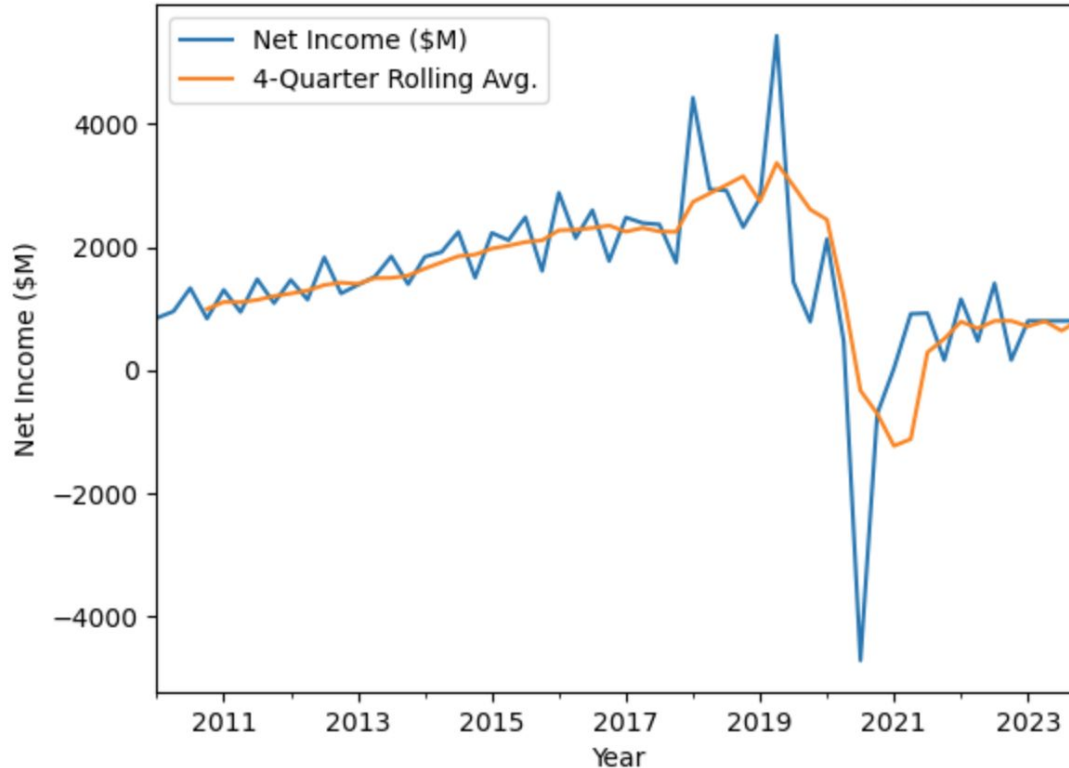
Forecasted Net Income for Y2023, Q1-4

Quarter_end	SMA4
2023-03-31	709.8125
2023-06-30	791.8750
2023-09-30	639.1875
2023-12-31	798.2500

* External factors like market conditions should be considered alongside the forecast for comprehensive financial planning.

Classification

Predicting Return Customers for Disney's Cruise Line

Modeling Process

Goal: Improve prediction of households likely to purchase follow-up voyage tickets after Disney Cruise using Logistic Regression and Random Forest models.

Data preprocessing: Handle missing values, encode categorical variables.

Classification

- **Logistic Regression Model**
  - Used to model probability of returning cruise customers.
  - Analyzes features like discounts, income, spending habits, & demographics.
  - Provides simplicity & interpretability for understanding feature impact.
- **Random Forest Model**
  - Handles nonlinear relationships & feature interactions.
  - Combines decision trees to enhance predictive accuracy.

Exploring the cruise returners dataset...
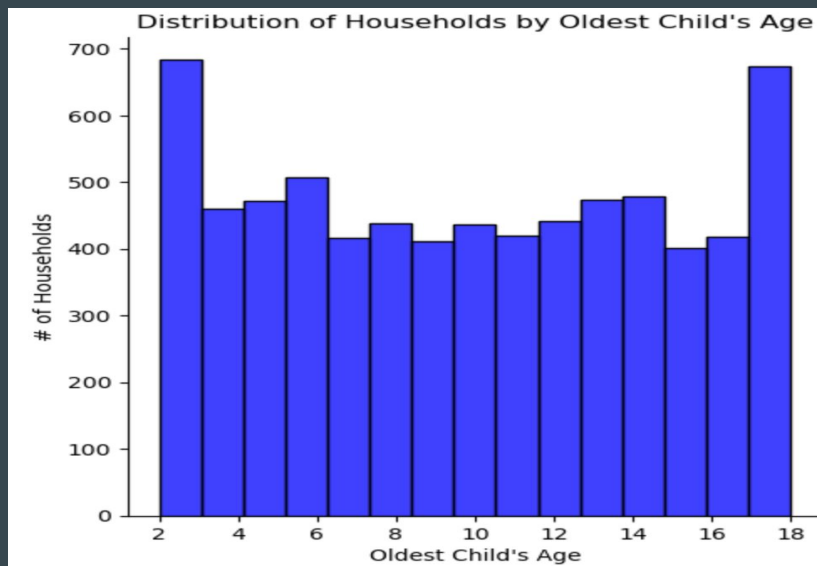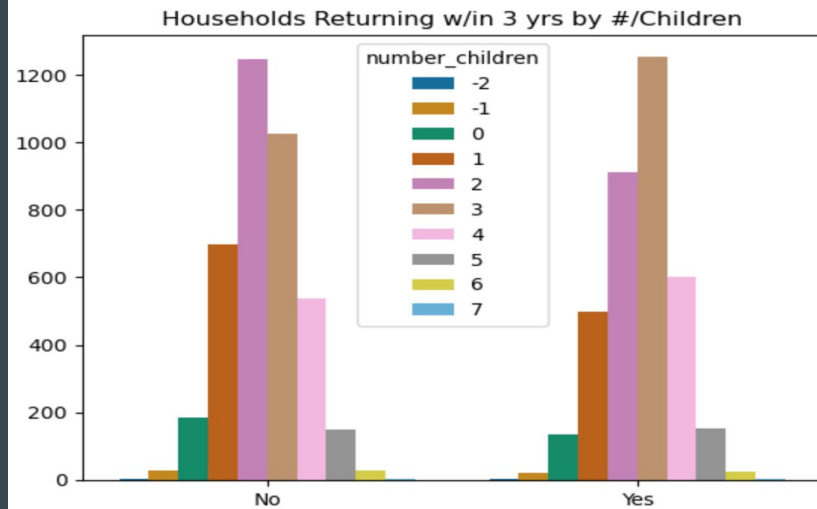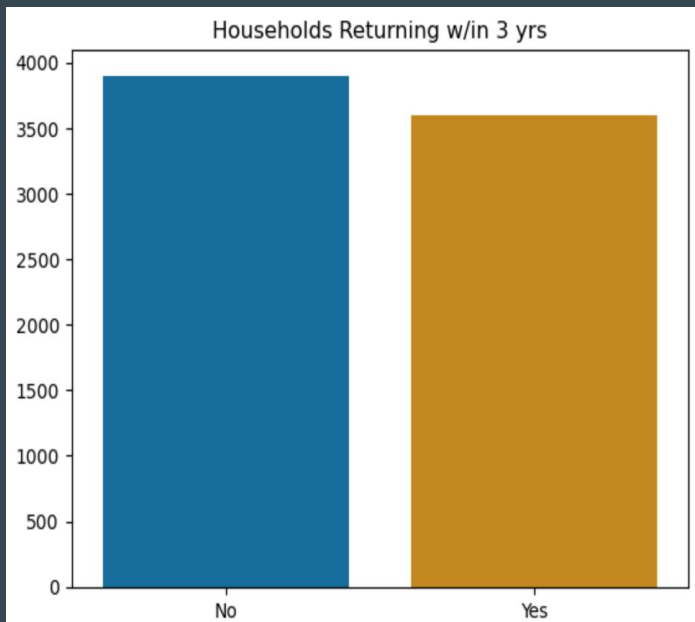
Model Output + Validation

Model **coefficients** reveal *influential factors* like oldest child age, distance to port, estimated income, etc.

```
Logistic Regression Coefficients:
                              Coefficient
discount_original             -3.133175e-05
coupon_received               -1.589110e-07
est_inc                        4.354165e-08
ticket_spend                   1.100762e-05
incidental_spending            5.716558e-06
park_package                  -3.778581e-05
res_to_port                   -2.813130e-03
number_children                5.915959e-05
oldest_child_age              -2.217219e-04
dis_plus_sub                   3.057570e-07
cabin_type_Concierge           2.525768e-06
cabin_type_Inside             -7.486047e-07
cabin_type_Oceanview          -1.102918e-06
cabin_type_Verandah           -4.056097e-07
season_of_year_Fall            1.405945e-07
season_of_year_Spring          3.232007e-06
season_of_year_Summer         -4.609847e-07
season_of_year_Winter         -2.642982e-06
cruise_theme_Marvel            3.608172e-07
cruise_theme_Princess/Frozen  -8.428398e-07
cruise_theme_Star Wars         4.047510e-07
embark_port_Barcelona         -5.160922e-06
embark_port_Copenhagen        -1.229208e-05
embark_port_New York          -1.027730e-05
embark_port_Port Canaveral    -1.323778e-06
embark_port_San Diego          7.116335e-06
embark_port_Vancouver          2.220638e-05
```
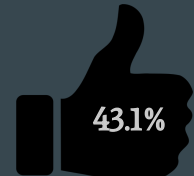
Example Prediction:

- Annual income = $100,000, # of children = 3, Oldest child age = 10y, Discount = No, Coupon = No, Distance to port = 350, Ticket spend = $3,109, Incidental spending = $1,889, Park package = Yes, Embark port = Vancouver

Prediction:

56.9%   vs.   43.1%

A/B Testing

Analyzing Popcorn Bucket Sales for Disney Theme Parks Division

Modeling Process

A/B Testing

Goal: Compare sales performance of different Disney popcorn bucket options.

- Isolated unique types of popcorn buckets & distinct theme park locations.

- Employed t-tests for each pair of popcorn bucket types or locations to calculate p-values.

- Set significance level at 0.05.

- *Null hypothesis:* mean difference in sales is zero (no statistical significance in difference).

- If **p-value < 0.05**, it indicates a *significant difference*, whereas if **p-value ≥ to 0.05**, there is *no significant difference*.

Model Output

```
Cinderella vs. Star_Wars: p-value=0.574 (not significant)
Cinderella vs. Mermaid: p-value=0.018 (significant)
Cinderella vs. MickeySuit: p-value=0.000 (significant)
Cinderella vs. MickeyEars: p-value=0.010 (significant)
Cinderella vs. RedWhite: p-value=0.000 (significant)
Star_Wars vs. Mermaid: p-value=0.017 (significant)
Star_Wars vs. MickeySuit: p-value=0.001 (significant)
Star_Wars vs. MickeyEars: p-value=0.049 (significant)
Star_Wars vs. RedWhite: p-value=0.000 (significant)
Mermaid vs. MickeySuit: p-value=0.000 (significant)
Mermaid vs. MickeyEars: p-value=0.000 (significant)
Mermaid vs. RedWhite: p-value=0.000 (significant)
MickeySuit vs. MickeyEars: p-value=0.063 (not significant)
MickeySuit vs. RedWhite: p-value=0.000 (significant)
MickeyEars vs. RedWhite: p-value=0.000 (significant)
```

* No significant differences found across Disney theme park locations

Implications:

- Focus on designs with consistently high sales (e.g., Star Wars, RedWhite).

- Reevaluate designs with lower performance (e.g., Cinderella).

- Implement uniform marketing strategies across all locations.
  - Emphasize popular designs.
  - Promote collector culture.
  - Ensure availability across all parks.

Use insights obtained in analysis to inform strategic decisions regarding the 4 P's of the marketing mix – product, price, place, & promotion...

- Explore movie performance metrics, including average gross revenues by genre and MPAA rating, through exploratory data analysis – Product.
- Know which customers are alike based on spending/purchase behaviors and other demographic attributes, which of those consumer segments to target, & how to target them – Promotion.
- Recommended proper amenities through conjoint analysis with survey data – Product & Price
  - Data provided how a single dollar invested can increase the customer experience rating by x amount.
  - Applied this analysis to maximize $250 investment per room per night for visitors on amenities.
- Inform 'placement' decisions in terms of whether to open certain theme-park related locations (e.g., Disney Hawaiian hotel, DisneyQuest) – Place.
- Optimize marketing plans and promotion strategies for Disney cruise business by understanding the impact of different features on predictor variables – Promotion.
- Forecast net income post-COVID-19 using a simple moving average approach – Price.
- Compare sales performance of different product options (e.g., popcorn buckets) through A/B testing to optimize marketing strategies – Promotion.