Roller Coaster Tycoon

Conjoint Analysis

Emili Khachatryan, Marianna Rodriguez, Jessie Calix, Julia Ruiz, Omar Shiqim, Naji Danial

Probability and Statistics for Data Management and Analysis

Rafif Srour

December 5th, 2023

Table of Contents

Abstract

To examine the emotional implications of roller coasters on individuals, it is crucial to conduct a proper analysis that evaluates how different characteristics of roller coasters affect the level of excitement experienced by the riders. The purpose of this project is to determine the roller coaster attributes that increase excitement among riders and to propose business strategies for designing roller coasters that effectively evoke a sense of amusement in people. For this purpose, a conjoint analysis will be performed on a data set containing information on user ratings across different attributes and the physical characteristics of numerous roller coasters. The results of the analysis indicated that the factors increasing the level of excitement were the *Number of drops* and the *Maximum speed* of the roller coaster. Conversely, the duration of the ride appeared to have a comparatively low impact.

Key Words:

Conjoint Analysis, Roller Coaster, Excitement, Linear Regression

Introduction

Today's dynamic market makes it essential for businesses to understand their customers. Companies must be adept at identifying and responding to the preferences and needs of their clients. To achieve this objective, market research and customer comprehension is completely necessary. Achieving a profound understanding of customers demands a comprehensive approach that exceeds basic demographic data. Successful businesses employ extensive multifaceted strategies to understand the intricacies of consumer behavior and preferences. By applying advanced data analysis techniques, companies can extract significant insights to enhance decision-making processes and improve sales by predictive consumer behavior.

In recent years, technology has advanced at a rapid pace. Individuals are increasingly more dependent on technological advancements as time progresses. Some positive and negative impacts emerge from this fact. The major negative effects that arise from the high reliance on technology are job displacement, digital inequality, environmental impact, and privacy issues, among others. However, many positive effects arise from technological advancement. Technology has allowed organizations to discover major breakthroughs across different industries and companies to increase their revenue through data analytics techniques.

Conjoint analysis is a statistical technique that is based on surveys and is used to study how individuals make decisions and how much weight they give to certain variables. This technique is used in a variety of industries, including research and marketing because it extracts consumer preferences using intricate models. A summary of conjoint analysis and related types of analysis is given by the Harvard Business School. "For example, an online store selling chocolate may find through conjoint analysis that its customers primarily value two features: Quality and the fact that a portion of each sale goes toward funding environmental sustainability efforts. The company can then use that information to send different messaging and appeal to each segment's specific value" (Harvard Business Review, 2020)[1]. It looks at how conjoint analysis is used in research and development, sales and marketing, and pricing. Conversely, Qualtrics[2] explores conjoint analysis and presents seven varieties of the technique. In addition to conjoint analysis, hierarchical Bayes analysis

---

[1]*What is Conjoint Analysis & How can you use it?: HBS Online* (2020) *Business Insights Blog*. Available at: https://online.hbs.edu/blog/post/what-is-conjoint-analysis (Accessed: 05 December 2023).

[2] Qualtrics. (2023, August 25). 9 common types of conjoint analysis and how to use them. [Webpage]. https://www.qualtrics.com/blog/conjoint-analysis/

is presented. Lastly, a piece from Westlaw Today[3] describes conjoint analysis and its uses, particularly concerning how customers make decisions. It also talks about the flaws in its presumptions. “In fact, as described in this article, conjoint analysis is founded on fundamentally inaccurate assumptions, namely that precise, stable preferences relatively accurately determine consumer choices, and that these preferences can be measured through tasking consumers to compare options comprised of a list of pre-specified attributes” (Westlaw Today, n.d).

Ultimately, conjoint analysis is still a powerful method for comprehending the complex world of consumer behavior. It determines consumer preferences by giving importance ratings to various qualities that comprise a product or service. Conjoint analysis finds applications within the roller coaster industry, where the type of analysis that is performed depends on the nature of the data and the objectives of that specific research. Although conducting this type of analysis has inherent limitations such as measuring emotional sensation in people, it remains a valuable technique in understanding and addressing consumer preferences.

Objectives:

1. Conduct a thorough investigation on conjoint analysis.
    - Discover the major applications of conjoint analysis across different industries.
2. Analyze the impact of different roller coaster attributes on rider excitement.
    - Discover the specific factors that increase the excitement experienced by the riders.
3. Propose various business strategies for the development of roller coasters aimed at increasing customer amusement.
    - Suggest roller coaster design implementations that provide an exciting and thrilling experience for their customers.

[3] Westlaw Today. (n.d.). The weak foundations of conjoint analysis: Methodological limitations for assessing consumer preferences and estimating damages in litigation settings. Secondary Sources. National. [Webpage]. Retrieved from https://today.westlaw.com/Document/Ica1fad58af8411ec9f24ec7b211d8087/View/FullText.html?bhcp=1&contextData=(sc.Default)&firstPage=true&transitionType=Default

Concept Development

The RollerCoaster Tycoon dataset contains information on in-game roller coasters, including ratings and physical attributes. We aim to perform a conjoint analysis to identify patterns in excitement, intensity, and nausea ratings based on factors like ride length, max speed, and ride time. This can help us find differences in between roller coasters, such as those with long rides and high speeds, or those with short rides and high intensity, providing insights for players, designers, and researchers. The data was collected through user surveys in which respondents rated different roller coaster rides based on these multiple attributes. The original data set contains 71 observations with 17 variables. Some of the variables are numerical and the others are categorical.

The primary purpose of the project is to understand the level of excitement generated by the different roller coasters and identify influencing factors. Conjoint analysis will be performed to determine the relative importance of different attributes and levels, as well as any interaction effect between the variables that affect roller coaster ratings. Conjoint analysis, however, is dependent on the presence of categories, so the following course of action was to create categories and dummy variables (three to four per each) from the relevant numerical variables and keep the other variables that were already in categorical form, including the dependent variable *Excitement level*.

The data preparation process was conducted in R Studio. The first step regarding the data cleaning process was checking for missing values and duplicate rows. The results indicated that there were no missing values or duplicate rows present in the data set, enabling further data preparation. Before creating categories within the variables, some downsides have to be taken into consideration: overfitting and weak model performance. Creating too many dummy variables will increase the chances of overfitting the model as there will be a small amount of observations per variable. Nevertheless, having categories is the essence of conjoint analysis. The number of categories created per numerical variable was determined by its range. Variables with higher ranges were assigned four categories, while variables with a smaller range were assigned three categories. Each category within a single variable encompasses an equal and continuous subset of the range to ensure an even distribution among levels. After deciding on the number of categories made for each variable, these were then transformed into dummy variables, and assigned as factors. The resulting dataset was stored and saved to be utilized for further analysis.

Results and Discussion

The conducted analysis delves into the exploration of the relationship between *Excitement level*, represented numerically, and six categorical variables expressed as dummy variables (*Nausea Level, Intensity Level, Maximum speed, Ride length (time), Amount of drops, Highest drop*). The choice of Ordinary Least Squares (OLS) regression as the analytical method is driven by the need to discern the optimal linear connection between *Excitement level* and categorical predictors. Ordinary Least Squares regression serves as the foundation of the analytical framework. By minimizing the sum of squared differences, OLS constructs the optimal linear relationship between the *Excitement level* (dependent variable) and categorical predictors. Residuals, representing the vertical distances between observed and predicted values, provide nuanced insights into this relationship.

To perform OLS analysis categorical variables have been efficiently transformed into dummy variables using the *get_dummies* method. This step ensures that categorical information is seamlessly incorporated into the regression model. Simultaneously, the *Excitement level*, maintained as a numerical dependent variable, facilitates a comprehensive examination of its variation across different factors.

The application of OLS to the cleaned data has produced a range of metrics and coefficients. These metrics serve as indicators of the model's performance and its ability to capture underlying patterns. The following sections will explore these metrics in detail, elucidating the significance of each predictor in influencing *Excitement level*. The breakdown of metrics will unveil insights into the impact of each categorical variable on customer excitement. This interpretative layer is crucial for translating statistical findings into actionable strategies, allowing for informed decision-making.

Results:

<u>R-Squared</u>: The model explains 57.2% of the variability in the *Excitement level* (dependent variable).
<u>Adjusted R-Squared</u>: Approximately 48.1% of the *Excitement level* variation is explained by the included variables, accounting for the number of predictors.
<u>F-statistic</u>: With a low p-value (< 0.001), indicating overall model significance, at least one independent variable significantly contributes to predicting *Excitement level*.

Dep. Variable:	excitement_lvl	R-squared:	0.572
Model:	OLS	Adj. R-squared:	0.481
Method:	Least Squares	F-statistic:	6.335
Date:	Fri, 01 Dec 2023	Prob (F-statistic):	5.79e-07
Time:	15:27:17	Log-Likelihood:	-69.598
No. Observations:	70	AIC:	165.2
Df Residuals:	57	BIC:	194.4
Df Model:	12		
Covariance Type:	nonrobust		

Table 1 Model evaluation

Each coefficient represents the change in the dependent variable for a one-unit change in the corresponding independent variable, holding others constant. Significant coefficients, such as *Intensity_lvl_high* (coefficient: 1.2589, p-value: 0.000), highlight statistically significant predictors. Visualization aids in identifying key variables for constructing captivating roller coasters.

	coef	std err	t	P>|t|	[0.025	0.975]
intensity_lvl_High	1.2589	0.159	7.902	0.000	0.940	1.578
intensity_lvl_Medium	1.3246	0.259	5.114	0.000	0.806	1.843
intensity_lvl_Very High	1.4031	0.225	6.235	0.000	0.952	1.854
nausea_lvl_High	1.4047	0.207	6.789	0.000	0.990	1.819
nausea_lvl_Low	1.0717	0.312	3.441	0.001	0.448	1.695
nausea_lvl_Medium	1.5102	0.156	9.673	0.000	1.198	1.823
max_speed_High	1.8916	0.364	5.203	0.000	1.164	2.620
max_speed_Low	0.7989	0.194	4.114	0.000	0.410	1.188
max_speed_Medium	1.2961	0.206	6.297	0.000	0.884	1.708
amt_time_ Very High	1.1496	0.434	2.649	0.010	0.281	2.019
amt_time_High	1.1944	0.240	4.981	0.000	0.714	1.675
amt_time_Low	1.0189	0.458	2.226	0.030	0.102	1.935
amt_time_Medium	0.6238	0.205	3.036	0.004	0.212	1.035
num_drops_High	2.0151	0.312	6.465	0.000	1.391	2.639
num_drops_Low	0.8101	0.182	4.443	0.000	0.445	1.175
num_drops_Medium	1.1615	0.179	6.504	0.000	0.804	1.519
high_drop_Low	0.8144	0.192	4.245	0.000	0.430	1.199
high_drop_Medium	1.2807	0.207	6.178	0.000	0.866	1.696
high_drop_Very High	1.8916	0.364	5.203	0.000	1.164	2.620

Table 2 Coefficients and their significance

Normality Tests (Omnibus, Jarque-Bera): P-values (0.794 and 0.734) suggest residuals may be normally distributed.
Durbin Watson Test: A value of 2.207 indicates moderate positive autocorrelation in residuals.
Multicollinearity Warning: A note highlights potential issues. Strong multicollinearity or a singular design matrix may affect coefficient stability and interpretability.

Omnibus:	0.461	Durbin-Watson:	2.207
Prob(Omnibus):	0.794	Jarque-Bera (JB):	0.619
Skew:	0.128	Prob(JB):	0.734
Kurtosis:	2.617	Cond. No.	3.28e+17

Table 3 Residuals evaluation

Due to the nature of Tycoon Roller Coasters' customer data, traditional assumptions like the normality of residuals are less applicable which is why the results from the tests are not completely reliable. Conjoint analysis was chosen for its robustness with diverse data types and resistance to non-normal distributions and it allows for meaningful analysis of emotional and preference data. It is tailored to individual-level preferences and eliminates the need for assumptions about joint distribution across the population. This characteristic enhances the method's flexibility, making it suitable for Roller Coasters' dataset where customer opinions and emotions are inherently subjective.

The coefficients are also displayed in Figure 1. It shows what effect the different variables had on the *Excitement level*, where a longer bar shows a bigger influence. Since the independent variables are encoded as dummy variables for each category the corresponding dummy and its weight can be seen as added points for the *Excitement level*. Not significant coefficients are marked in red so it can be seen that every coefficient is statistically significant.

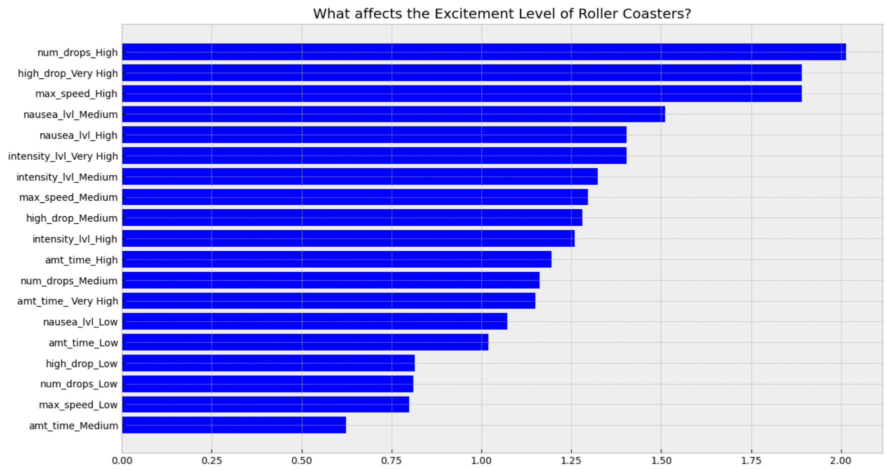

Figure 1 Effect of the coefficients on Excitement level

When looking at the *Number of drops* and the *Highest drop* a high respectively very high value together with a high *Maximum speed* value had the biggest effect on increasing the *Excitement level*. Next comes the *Nausea level*. Here a high value also increases the *Excitement level*, here it would be expected that it is the opposite case. This might be due to the correlation with the other already mentioned variables, which was also mentioned in the multicollinearity warning from the model. In general, it can be seen in Figure 1 that a low level in one of the variables has also a lower increase in *Excitement level*, than a higher level.

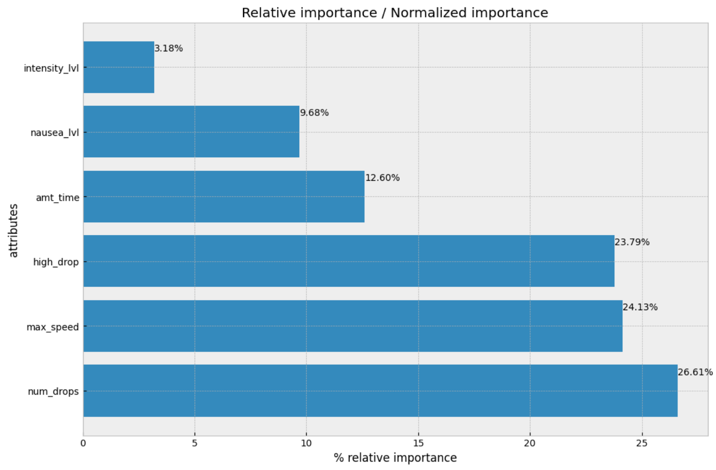

Figure 2 Relative importance of the different categories

Next to the type of level that influences the dependent variable, the variable type that has the biggest influence is in a conjoint analysis very important since it can give information about the preferences of the customer and tell where to focus. This is displayed in Figure 2. It shows the relative importance of the different variables, which is calculated by dividing individual importance by the total importance. It can be seen that the *Number of drops*, *Maximum speed*, and *Highest drop* have the biggest influence on the dependent variable.

Conclusion and Recommendations

In analyzing the roller coaster data set, the goal was to determine the variables that significantly impact excitement ratings to design a perfect roller coaster for an amusement park. After analyzing the results, the *Number of drops* and the *Maximum speed* of the roller coaster proved to be the factors that increased the *Excitement level* the most. However, given the accuracy of the model, it can not be stated that there is any practical significance in the model at the moment despite the statistical significance of some of the variables. Most of the variables do not exceed the necessary p-value threshold while others may be very subjective when it comes to interpretation (i.e. *Intensity Level* and *Nausea level* may vary from person to person). Essentially, the suspected primary cause of low model predictive power is the data collection strategy.

To increase the accuracy of the model, consideration could be given to redesigning the user survey or altering the approach to data collection. For instance, instead of asking people about their feelings verbally, collecting objective biometric data (i.e. heartbeat rate) may prove more effective. In the analysis that involves user opinions and subjective thoughts, data collection is the most crucial aspect of the entire analysis. The proposed initiative aims to improve the model and, in the future, design a perfect roller coaster or a set of roller coasters that would satisfy the needs of every customer group. The ultimate goal is to ensure that visitors leave the amusement park at their highest *Excitement level* with unforgettable emotions.

Bibliography

- Qualtrics. (2023, August 25). 9 common types of conjoint analysis and how to use them. [Webpage]. https://www.qualtrics.com/blog/conjoint-analysis/

- Bhaskaran, V. (2023, October 26). Conjoint Analysis: Definition, example, types and model. QuestionPro. [Webpage]. https://www.questionpro.com/blog/conjoint-analysis/

- Westlaw Today. (n.d.). The weak foundations of conjoint analysis: Methodological limitations for assessing consumer preferences and estimating damages in litigation settings. Secondary Sources. National. [Webpage]. Retrieved from https://today.westlaw.com/Document/Ica1fad58af8411ec9f24ec7b211d8087/View/FullText.html?bhcp=1&contextData=(sc.Default)&firstPage=true&transitionType=Default

- *What is Conjoint Analysis & How can you use it?: HBS Online* (2020) *Business Insights Blog*. Available at: https://online.hbs.edu/blog/post/what-is-conjoint-analysis (Accessed: 05 December 2023).

- Author, N. (2023, June 2). Twists, turns, thrills and spills: the physics of rollercoasters – Physics World. Physics World. https://physicsworld.com/a/twists-turns-thrills-and-spills-the-physics-of-rollercoasters/

Appendix

Data Set Source:
GitHub

Data Set Link:
Google Sheets

Data Set Overview:

0	Barony Bridge	Bobsleigh Coast	0	5.49	High	5.71	High	4.86	Medium	39	12	63	1496
0	Barony Bridge	Dinghy Slide	0	4.83	Medium	5.86	High	3.5	Medium	37	11	48	1079
0	Barony Bridge	Wild Mouse	0	7.64	High	7.86	Very High	4.87	Medium	31	11	70	1591
0	Barony Bridge	Wooden Roller C	0	7.69	High	7.92	Very High	4.75	Medium	47	15	79	2401
1	Forest Frontiers	Junior Roller Co	0	5.17	High	5.54	High	3.81	Medium	34	13	51	1279
2	Haunted Harbou	Side-Friction Rol	0	5.68	High	6.18	High	3.6	Medium	32	10	75	1466
2	Haunted Harbou	Wooden Roller C	0	7.76	Very High	7.62	High	4.68	Medium	45	17	63	2077
2	Haunted Harbou	Wooden Wild Mo	0	7.19	High	7.88	Very High	4.69	Medium	30	12	53	1246
3	Mystic Mountain	Stand Up Roller	1	5.36	High	8.66	Very High	5.97	High	41	14	48	1358
3	Mystic Mountain	Wooden Roller C	1	6.33	High	7.24	High	4.56	Medium	41	9	74	1378
4	Pacific Pyramids	Compact Inverte	1	6.05	High	7.64	High	6.37	High	38	9	73	1328
4	Pacific Pyramids	Looping Roller C	1	5.76	High	5.44	High	2.49	Low	36	13	50	1335
4	Pacific Pyramids	Stand Up Roller	1	2.72	Medium	5.29	High	4.25	Medium	40	12	89	2152
4	Pacific Pyramids	Vertical Drop Co	1	3.11	Medium	6.91	High	3.98	Medium	45	13	68	1762
5	Mel's World	Bobsleigh Coast	0	6.32	High	5.47	High	4.71	Medium	35	15	91	2700
5	Mel's World	Inverted Roller C	0	7.37	High	9.57	Very High	7.63	High	51	20	65	2595
5	Mel's World	Spinning Wild M	0	6.32	High	6.69	High	5.42	High	30	1	55	1246
5	Mel's World	Suspended Swir	0	6.67	High	6.54	High	7.13	High	43	13	77	1972
5	Mel's World	Dinghy Slide	1	5.28	High	5.16	High	3.3	Medium	32	8	62	981
5	Mel's World	Looping Roller C	1	6.42	High	7.39	High	4.2	Medium	38	12	76	1778
5	Mel's World	Vertical Drop Co	1	6.91	High	6.59	High	3.32	Medium	50	12	56	1581
6	Paradise Pier	Dinghy Slide	1	5.21	High	5.41	High	3.44	Medium	32	9	72	1322
6	Paradise Pier	Looping Roller C	1	5.7	High	5.03	Medium	2.48	Low	37	11	69	1522
6	Paradise Pier	Stand Up Roller	1	2.55	Low	3.59	Medium	2.61	Medium	34	9	58	1036
7	Gentle Glen	Junior Roller Co	0	5.26	High	5.29	High	3.6	Medium	34	12	60	1420
7	Gentle Glen	Dinghy Slide	1	2.46	Low	2.56	Medium	1.6	Low	29	9	52	922
7	Gentle Glen	Mini Roller Coas	1	5.75	High	5.05	Medium	4.03	Medium	36	11	84	1873
7	Gentle Glen	Virginia Reel	1	5.07	High	5.35	High	5.69	High	32	10	52	1030
8	Jolly Jungle	Heartline Twiste	0	3.12	Medium	6.6	High	4.61	Medium	39	10	33	620