To: Big City Bikes CEO: Alexandra Wheeler
From: Kevin Ortiz - Larson's 8AM Business Analytics
Subject: Business Intelligence at Big City Bikes
Date: December 3rd, 2023

Based on the available data from the last 28 months, I recommend focusing on increasing the usage of classic bikes during the summer & warmer periods and focusing on regions with higher total populations.[i] While countless reiterations of models were created to test different economic features predictive of crime trends, the final model was more restricted to help with predictive accuracy[ii], I believe this recommendation would be the best course of action.

The general approach in this conclusion went as follows: **1) Data Cleansing [iii]2) Post spatial Analysis Data Filtering 3) Refining revenue as target 4) Predictive Modeling Accuracy**.[iv] Several different factors were tested using the historical trip data including: ride duration, revenue trends, bike type/membership trends, station/county/zip code popularity, and socio-economic factors.[v] The two external data sources, zip code data and census data, were used to help with the all-encompassing study.[vi]

Data Analysis and Overall Rental Trends Over the Past 28 Months:

1. In analyzing the trends, comparing the bike types showed favorability toward the classic bike types in the higher performing stations more than other types.
2. Station location was a huge factor showing favorability in zip codes and counties with higher populations.
3. Time periods with warmer conditions like summer months showed higher performance.

Crime Trend Correlation: Regarding the correlations between crime trends and rental trends, the evidence revealed insufficient evidence correlation between rentals and crimes. Exhaustive regressions were performed using secondary census and zip code data merged with the primary data set, however accuracy never passed a reasonable threshold to provide recommendations.

Actionable Recommendations:

1. Promotion/Increase of Classic Bikes for both cities: Launching a promotion and increasing the number of bikes to incentivize the use of classic bikes, which is seeing a consistently growing demand, and higher correlation in revenue than the other bike types.

2. Focus on Warmer Periods: Offer incentives or collaborate with tourist attractions to promote and educate possible riders of bike rentals as a means of transportation. Resources are better spent when there is the largest demand for bike rentals.

3. Focus on Popular Locations: There are no metrics indicating stock-outs in popular stations which may be a huge factor in preventing the increase of rentals. A focus on these popular areas by means of maintenance, inventory and promotion would have a huge effect on increasing rentals.

To: Big City Bikes CEO: Alexandra Wheeler
From: Kevin Ortiz - Larson's 8AM Business Analytics
Subject: AI Supplication
Date: December 3rd, 2023

Ideation: AI was used in the very inception of the project to come up with insights on how to kickstart the project and outlining the steps that should be taken. The tool was able to break down the project into manageable steps that should be taken and creating a comprehensive roadmap. AI streamlined the project's commencement, allowing for a fail forward approach and expediting the definition process.

Data Analysis: Harnessing AI's power of text data analysis to analyze the text associated with the project and getting a better understanding of the initial readings was crucial. The tool was particularly helpful in analyzing the census data, swiftly gathering relevant information, and making data driven recommendations on what datasets would be more helpful in finding the crime trends.

Data Cleaning: AI significantly sped the cleaning process, offering an efficient alternative to tools like Alteryx. The tool powered the data cleaning process, enhancing the accuracy and quality while saving time and resources.

Exploring New Analysis Approaches: In the realm of data analysis and other business scenarios, there are often situations where we need to venture into attempting new methodologies that haven't been explored. One such instance in the project was to find correlations between the data through a regression analysis, aiming to uncover positive correlations between the variables. While cases like these may be hard in identifying the best strategy, the use of AI made it seamless and easy.

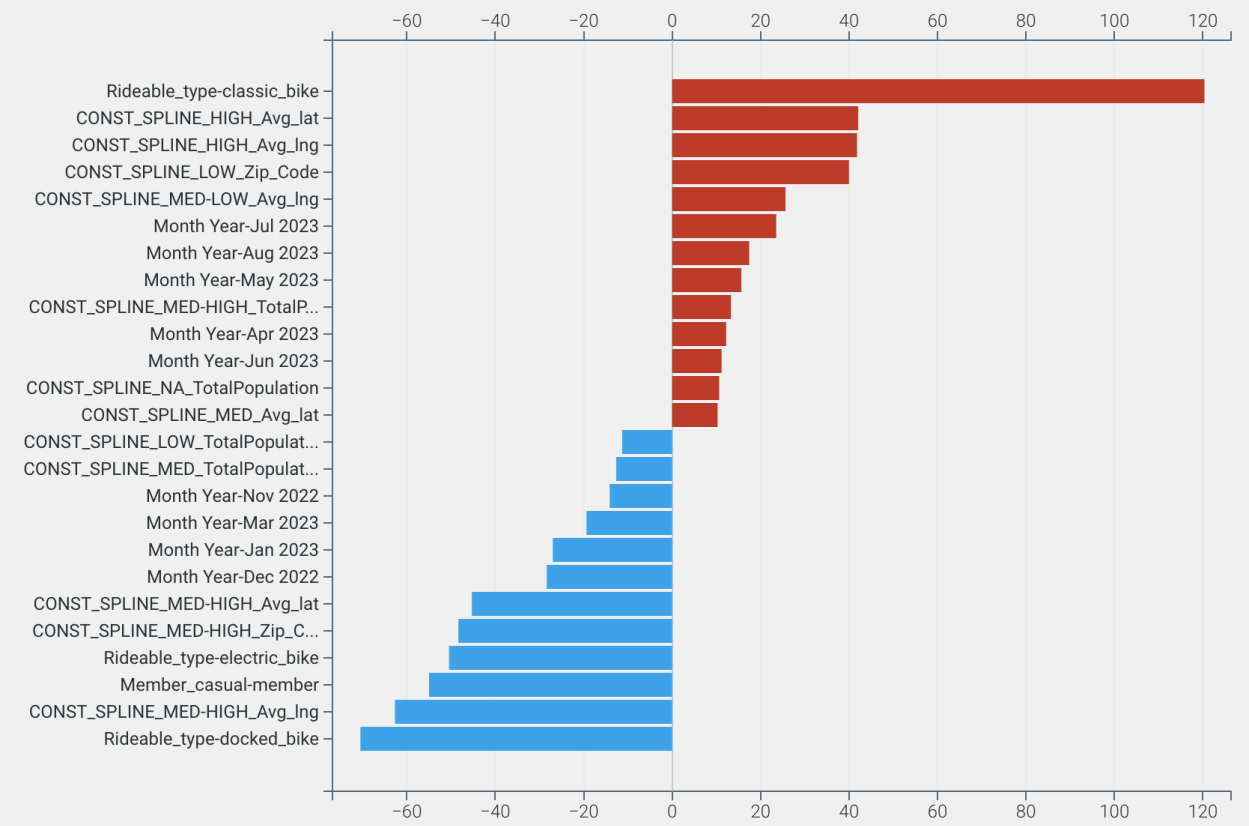
Exploring Different Use Cases: Throughout the project there were many instances where the use of a new tool and techniques were unfamiliar. In these moments of experimentation, AI emerged as a guiding light making faster routs or highlighting new tools that haven't been considered.

Future Implications: The potential for AI is boundless, there are countless of use cases from automating time-consuming processes and optimizing the productivity of high performing employees. Looking towards the future of the organization the implementation of AI is integral for offering continued opportunities in efficiency, innovation, and impact.

Conclusion: The seamless integration of Artificial Intelligence, from Idea generation to data analysis, has not only boosted productivity but also catalyzed innovation. AI's role in streamlining processes, uncovering hidden insights, and fostering innovative problem-solving approaches signifies the transformative force within the organization. Artificial Intelligence not only helped with efficiency and saving resources for this project but also fostered innovative approaches to problem-solving and uncovering hidden insights. The possibilities of AI are boundless, with endless opportunities for productivity innovation and impact for the betterment of this organization and the people it serves.

Appendix

[i] **Model Insights**



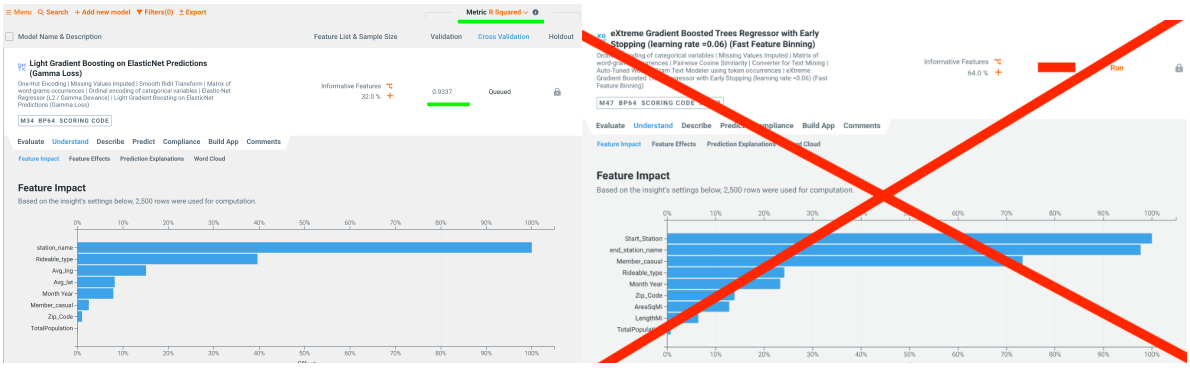
During the analysis, the model insights tool played a critical role. The tool played an instrumental role in making the predictive models faster to understand. The interplay between the features and their effects on the bike rental trends played a big part in looking over the distribution and its impact on the different characteristics within the model.

The first models developed had a skewed and abnormal skewed distribution. Model insights played a significant role in quickly visually analyzing each model to ensure an accurate distribution of descriptive features.

> Seasonal variations**:** There was a significant increase in summer rentals, which will be crucial for strategic planning.
>
> Demographic influence**:** The total population was a significant consideration in the overall model, highlighting its importance as a feature.

Additionally, longitude and latitude were left as features because of the potential correlations that could be found within specific regions or directions of the city.

[ii] **Comprehensive Iterative Modeling for Optimizing Rental and Revenue Predictions**



There was an exhaustive number of different iterations of models before choosing one. The process involved creating and validating many varying features, including their effects on the **primary target: revenue**.

A significant aspect of this process was balancing the amount of feature impacts and the accuracy of the model. Many different models fell short in accuracy despite highlighting important socioeconomic factors that would have been an excellent addition to the models. Meanwhile, other models had high accuracy but needed more crucial features in making practical recommendations. A large amount of experimenting was required not only to find a model with high accuracy but also to understand the overall nature of the different features that could be involved in making a recommendation. The final features that were selected were features that were not only practical but also accurate.

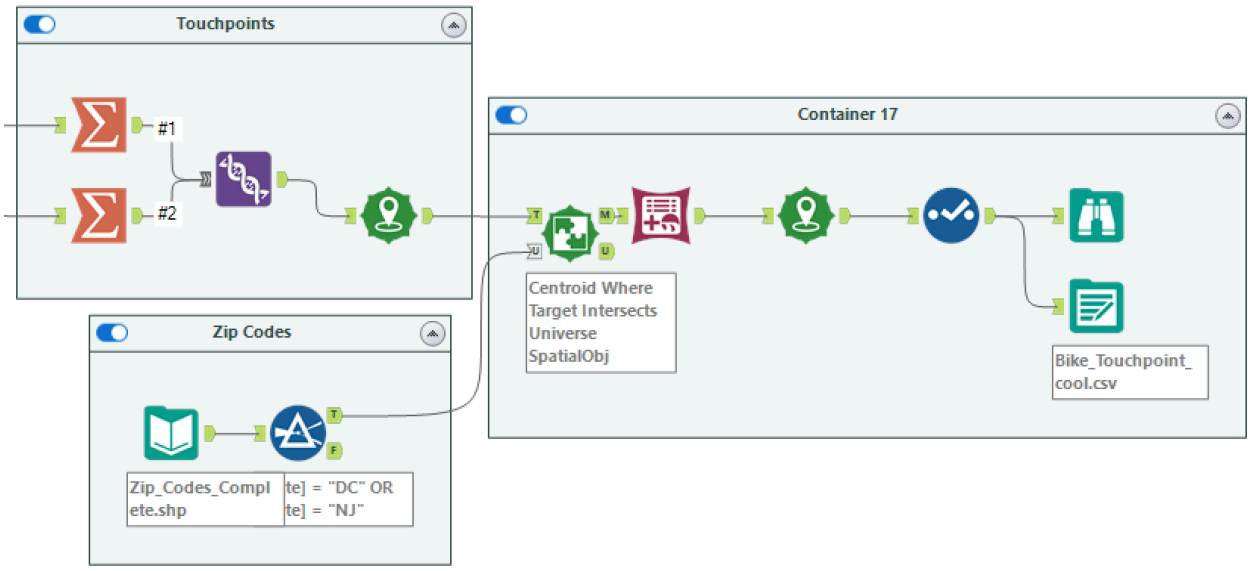
[iii] **Data Cleaning and Preparation Enhance Analysis in Project two**

The data preparation stage was a significant improvement. This phase was met with a complete overhaul of the initial dataset and workflows to make a dataset for Data Robot and increase rentals.

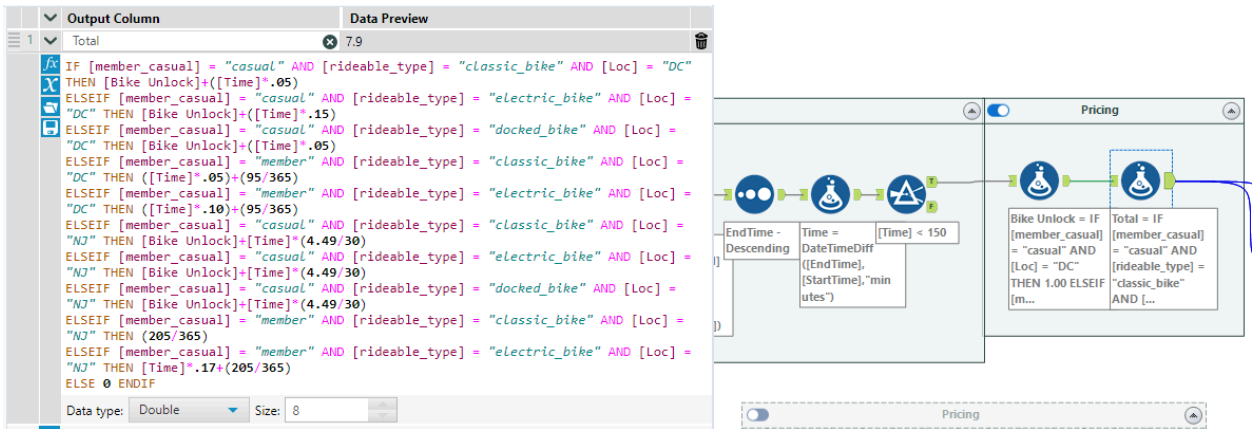
Key steps:

1. Data Pruning - After a few attempts with the original analysis, the original dataset and workflow were deemed unusable. As a result, much of the original workflow and data was not used, but the critical insight was used as a compass for this new attempt.
2. Integration of Zip Code Data – To help with the amount of possible information provided in the recommendation, zip code data was integrated. External data was combined to filter out unlikely data like several rental stations in highly populated areas such as New

York.



3. Focus on Pricing and Revenue Accuracy – A significant reason for a complete a workflow redesign was the accuracy in the pricing and revenue. This process involved calculating ride times for each ride and pairing them based on the different costs on the company's websites.



4. Spatial analysis preparation - The data was then treated more to accurately pair zip code and census data for each finished ride.

[iv] **Different Model Types**

Model Comparison

Model	R Squared (Validation)	R Squared (Cross-validation)	Gini Norm (Validation)	Gini Norm (Cross-validation)	Prediction Time(s)
Light Gradient Boosting on ElasticNet ... M34 | BP64 | 32% Sample Size | Informative F...	0.9337	0.9312	0.9791	0.9801	0.74
Light Gradient Boosted Trees Regresso... M37 | BP63 | 32% Sample Size | Informative F...	0.9287	0.9287	0.9800	0.9801	1.03

Making the final recommendation requires leveraging various iterations of models, which help in understanding and choosing to make the final predictive model.

Key Models and Their Contributions

1. Regression Regressor – The insights from this model were critical in the preliminary features selected in the data.
2. Random Forest Regressor – Compared to the regression regressor, this model provided more profound insights into the relationship between the features and their importance. The most helpful aspect was its highlighting of nuances in feature interactions and the collective combined impact on the model's accuracy.

The final selection for the model was based primarily on accuracy over speed, given the sufficient time available to make a proper analysis. Each model offers different perspectives on accuracy and efficiency; the final decision was based more on the requirements and restrictions.

[v] **Rigorous Feature Selection for optimized Rental and Revenue Prediction**

Feature Name	Data Quality	Index	Importance	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max
Sum_Total		6	Target	Numeric	18,064	0	5.95	18.30	2.81	-4.70	3,286
Start_Station		4		Text	767	0					
end_station_name		5		Text	470	0					
Zip_Code		7		Numeric	39	0	18,457	4,193	20,005	7,030	20,566
LengthMi		9		Numeric	39	0	8.79	5.04	7.93	0.32	21.61
AreaSqMi		8		Numeric	39	0	2.29	1.71	2.23	0.00	6.50
Member_casual		1		Categorical	2	0					
TotalPopulation		10		Numeric	11	493,364	18,033	10,306	15,173	2,790	65,880
Month Year		3		Categorical	12	0					
Rideable_type		2		Categorical	3	0					

The selection of the features for the predictive model resulted from an exhaustive series of iterations and analyses. The meticulous process was essential in identifying the most impactful features for predicting bike rental trends and improving revenue.

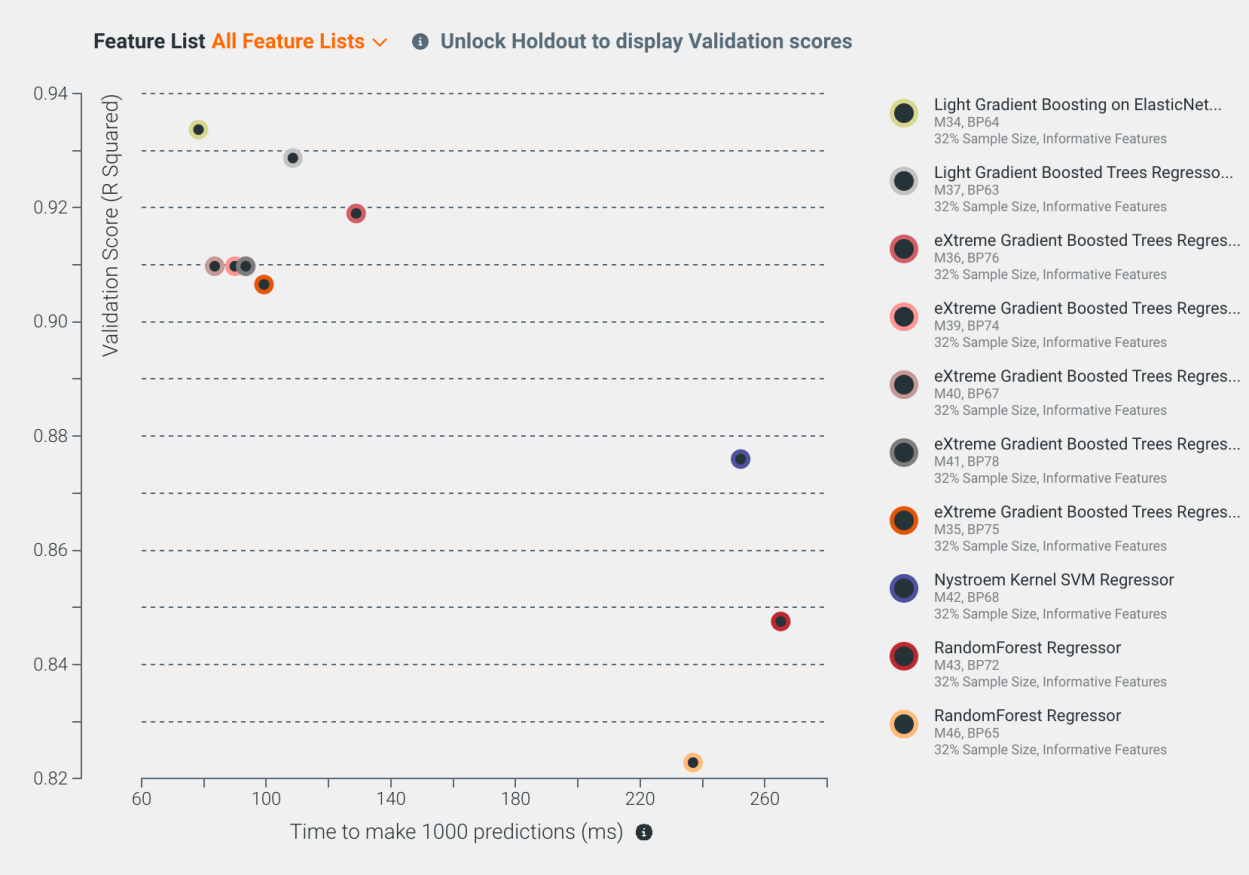
Final List of Selected Features

1. Sum_Total: Calculated by aggregating ride lengths for each ride ID categorized by rider type (member or casual), pricing data was later found on both companies' websites. This aggregation was further broken down by station, zip code, and county.
2. Start_ Station: Identifies the starting point of each ride.
3. End_Station_Name: Marks the endpoint of each ride.
4. Zip_Code: Sourced externally, providing location data for each station and additional demographic insights.
5. Length of Zip Code: Helps understand the geographical size of each zip code area.
6. AreaSquareMi: Similar to the previous feature, this gives an understanding of the scale and size of each zip code area.
7. Member_Casual: Categorizes Riders as members or casual users

8. TotalPopulation: Derived from external census data, offering insights into the population density within each zip code.
9. Month Year: Used to analyze seasonality and identify trends over time.
10. Rideable_Type: Details the type of bike used, providing insights into user preferences.

Each feature was carefully chosen to minimize data skew and remove variables that could affect the analysis. The goal was to ensure that every selected feature contributed to the model's ability to predict rental trends and revenue.

[vi] Model Selection and Accuracy



Light Gradient Boosting on Elastic Net Prediction

One of the first models employed in this comprehensive analysis was the light gradient boosting on elastic net prediction. The model was selected for its superior validation performance and overall effectiveness in the predictive analysis.

- Validation and accuracy - This model did not match the specifications of the other models, but its overall selection was strategic, considering ample time was available for conducting this analysis. The slight trade-off in performance was acceptable in this scenario, considering the time available for delivery and the accuracy needed to make such a big recommendation.
- Performance Considerations – This model did not match the specifications of the other models, but its overall selection was strategic, considering ample time was available for

conducting this analysis. The slight trade-off in performance was acceptable in this scenario, considering the time available for delivery and the accuracy needed to make such a big recommendation.

- Data utilization - This data underwent extensive cleaning and modification to ensure optimal feature representation to avoid over-sampling issues.

Additional Sources used for analysis and explanation

https://downloads.alteryx.com/data.htm
https://datablends.us/2017/07/31/a-useful-usa-zip-code-shapefile-for-tableau-and-alteryx/
https://capitalbikeshare.com/system-data
https://citibikenyc.com/system-data