Zen City's Journey through London's bike rental data - Roadmap

Over the past two weeks, we've been immersed in understanding the intricacies of "Zen City" bike rental business and the vibrant urban landscape of Zen City. This document serves as your roadmap, providing insight into the journey we undertook, the conclusions we reached, and the steps we took to enhance our services.

Before delving into the project's details, we took a moment to step back and fully grasp the essence of our business. We recognized the importance of comprehending the bigger picture before diving into the data. This initial contemplation helped us align our analytical efforts with the core objectives of Zen City.

Our Approach: Unveiling the Hidden Insights

In our pursuit of comprehensive analysis, we were motivated by a curiosity to explore beyond the surface. Instead of merely scratching the data's surface, we aimed to unravel the hidden gems nestled within. Our focus was not just on the routine or mundane, but on identifying novel angles that could offer valuable insights.

We decided to venture into uncharted territory by formulating business questions that stretched beyond the ordinary. These questions were meticulously designed to tap into patterns, preferences, and behaviors that might be concealed within the data. By doing so, we aimed to uncover the unique aspects that set Zen City's bike rental service apart.

Gaining Deeper Understanding: Patterns, Preferences, and Usage

Patterns were a central theme in our quest for insights. We sought to decipher the temporal rhythms and seasonal trends that shaped bike rentals. By understanding when our services experienced peaks and troughs, we could optimize our resources to better serve our users.

Client preferences emerged as another pivotal facet. We delved into the data to discern popular routes and station pairings, aiming to cater to the unique demands of our clientele. These preferences were integral in fine-tuning our station placements and product features.

Lastly, end-user usage patterns offered a glimpse into the daily lives of our customers. We analyzed ride frequencies, durations, and geographic clusters to anticipate user behavior. This information empowered us to enhance user experiences and tailor our services accordingly.

Embarking on Your Data Journey

This roadmap will guide you through the chapters that follow, each a stepping stone in our data-driven expedition. From data exploration to hypothesis formulation, from data cleaning to analysis and visualization, and ultimately to predictive insights, you'll navigate through each phase with a clear sense of purpose.

As you traverse this journey, remember that your role extends beyond a data analyst—you are a critical contributor to Zen City's mission of eco-friendly commuting, seamless user experiences, and data-driven excellence. Let's dive in, explore, and unveil the remarkable stories that our data holds.

Chapter 1: Enhancing Bike Rental Efficiency and Customer Experience

Defining the Goal: A Data-Driven Approach to Optimize Business Operations

Our journey begins with a clear mission: to optimize Zen City's bike rental operations for increased efficiency and to enhance the overall customer experience. The specific target is to drive more bike rentals during Q2 2021 (April to June) by fine-tuning station placements and product features.
While Zen City has strategically positioned bike stations, the company recognizes that there is room for improvement. To achieve this goal, we embarked on a data-driven analysis.

Setting the Stage: The Power of Data in Decision-Making

Before diving into data exploration, we realized the importance of understanding the business landscape and its challenges. The existing station placements, while strategically positioned, might not be optimal for maximum efficiency. By leveraging our data, we aimed to identify underutilized and overcrowded stations to make informed decisions that would positively impact the business.

Crafting Relevant Business Questions: Uncovering Key Insights

Our focus was on crafting meaningful questions that would help us uncover hidden insights within the data. The questions we devised were aimed at understanding customer behavior, preferences, and utilization patterns:

Ranking Stations for Efficiency: To optimize station placements, we posed the question: What are the best and worst performing starting stations, and what distinguishes them? This approach would allow us to pinpoint successful station setups and identify areas for improvement.

Maximizing Revenue by Time: We sought to answer the question: What are the best days of the week and hours to maximize revenue? This inquiry would enable us to align our resources with peak demand periods and better serve our customers.

Identifying Popular Ride Pairs: To enhance the user experience, we asked: What are the most popular ride pairs, and where do our customers prefer to ride? This exploration would help us enhance station placements and cater to user preferences.

Addressing Overcrowded Stations: By investigating the question of identifying overcrowded stations, we aimed to alleviate potential bottlenecks and ensure smoother operations.

Our Roadmap for Analysis: An Iterative Process

With our questions in hand, we proceeded to the data exploration phase. We sought to uncover temporal patterns, detect seasonal trends, identify outliers or anomalies, calculate average distances between stations, and explore the formation of spatial clusters of bike stations. This initial exploration guided us in understanding the data landscape and preparing for more in-depth analysis.

The analysis we performed was driven not only by data-driven insights but also by domain knowledge. By combining our analytical skills with a deep understanding of the business, we aimed to drive impactful outcomes.
In the following chapters, we will dive deeper into the specifics of each question, exploring data cleaning, analysis, visualization, and eventually predictive insights. With each step, we will unravel the stories hidden within the data and contribute to Zen City's mission of fostering eco-friendly commuting and enriching user experiences.

screenshot of the first business Q we wrote

Q1
What are the busiest days and times - check UTC and london.
Morning / Night - check the best station for each day and time.

Q2
Best pairs - can push coupons for usage between these pairs.
האם תחנות מאבדות מהערך שלכם לאורך

Q3
Start station best days and hours - if its full all day long please open new stations nearby.

Q4
Duration per Hour with weekdays - push for longer rides when ride time is short with coupons to make them longer.

Q5
Does old bike ride length get shorter or longer?

Q6
Calculate the longest rides with geo points and try to find the best stations to make better routes for sale.

Q7
Calculate the rush hour by rides per hour and know if you need to add bikes to the circle

Chapter 2: Unveiling Insights through Data Exploration

In this phase of our analysis, we dived into the heart of our data: the cycle_hire_new and cycle_stations_pro tables. These tables held the key to unlocking the patterns, preferences, and user behaviors that we sought to understand. Our data exploration aimed to derive meaningful insights that would fuel our efforts to optimize Zen City's bike rental operations.

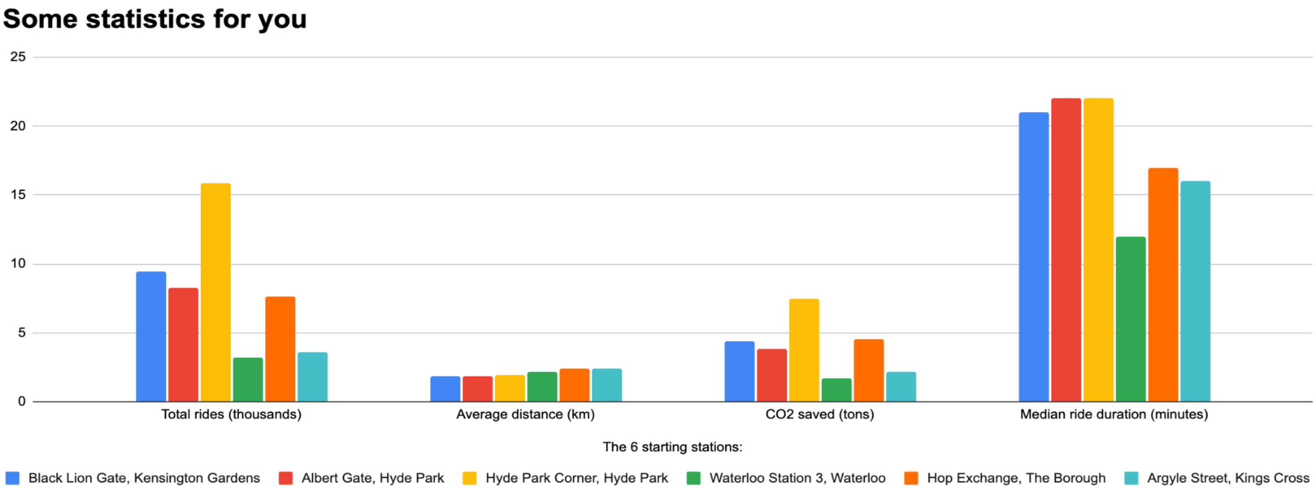
The cycle_hire_new table, containing information about each bike rental transaction, was a treasure trove of user interactions. We began by investigating temporal patterns and seasonal trends, meticulously charting how the demand for bike rentals fluctuated across different days, weeks, and months. By delving into this temporal dimension, we sought to uncover peak periods of demand and identify patterns that could guide resource allocation.

Simultaneously, we examined the cycle_stations_pro table, which provided us with crucial information about each bike station's location and capacity. Armed with this data, we aimed to identify spatial clusters of bike stations, allowing us to better understand the geographical preferences of our users. The relationship between station locations and user behaviors was paramount to our analysis, as it would inform station placement adjustments to cater to user demands more effectively.

As we embarked on data exploration, we remained attuned to potential outliers or anomalies that could impact our analysis. By combining our curiosity-driven approach with the nuances of the business, we were primed to unravel deeper layers of insight that would ultimately drive our efforts to enhance the bike rental experience in Zen City. In the subsequent chapters, we will delve further into the outcomes of our exploration, illuminating the correlations, trends, and stories hidden within the data.

Row	avg_distance_km	median_trip_distance	min_distance_km	max_distance_km	avg_duration_minute	median_duration_mi	min_duration_minute	max_duration_minut
1	2.02	1.829	0.0	12.96	28.79	20.0	1.0	328.0

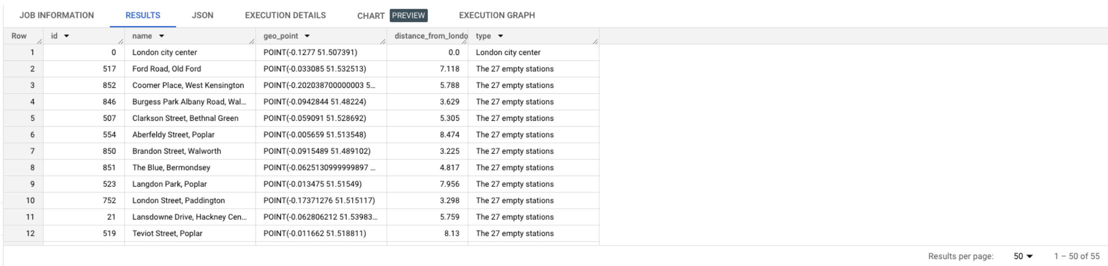
Row	starting_name	total_rides_per_static	AVG_AIRlength_dista	total_CO2_saved_by_	median_duration_mi
1	Black Lion Gate, Kensington Ga...	9.435	1.85	4.36	21.0
2	Albert Gate, Hyde Park	8.283	1.87	3.86	22.0
3	Hyde Park Corner, Hyde Park	15.845	1.9	7.51	22.0
4	Waterloo Station 3, Waterloo	3.21	2.15	1.72	12.0
5	Hop Exchange, The Borough	7.613	2.39	4.52	17.0
6	Argyle Street, Kings Cross	3.616	2.39	2.15	16.0



Chapter 3: Crafting Hypotheses for Data-Driven Success

Having delved into the data and gaining a profound understanding of the intricacies of Zen City's bike rental ecosystem, we are now ready to embark on the phase of formulating hypotheses and framing business-critical questions. As we peer into the data landscape we've explored, a series of inquiries arise that have the potential to reshape Zen City's strategies and foster significant growth.

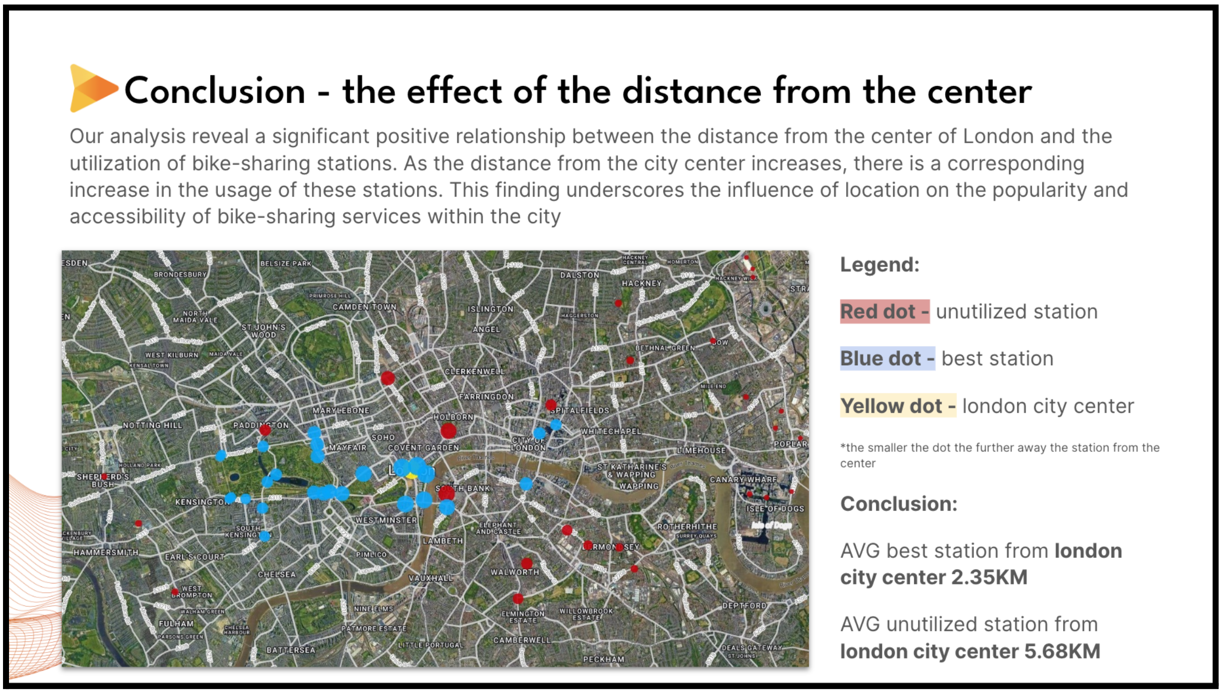
One of our key hypotheses revolves around station efficiency. By ranking the best and worst-performing stations and discerning the factors that set them apart, we aim to uncover insights that could inform station placements and elevate the overall customer experience. The understanding of why certain stations thrive while others struggle is fundamental to optimizing our network for maximum efficiency.



JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

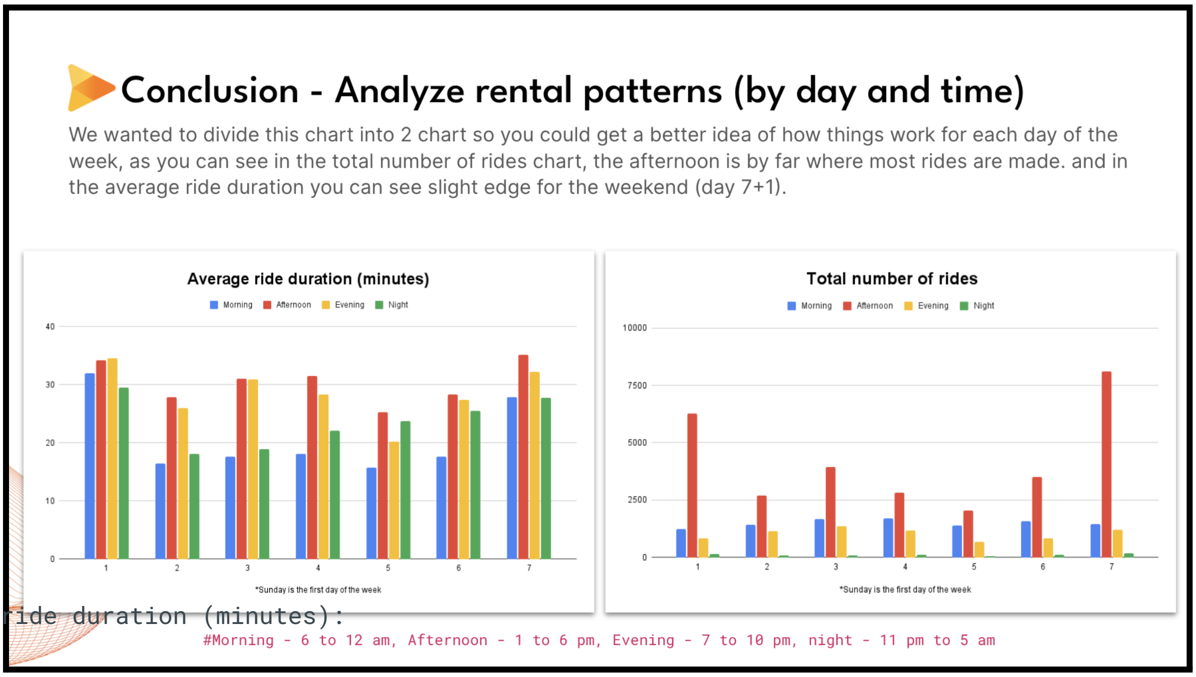
Row	id	name	geo_point	distance_from_londo	type
1	0	London city center	POINT(-0.1277 51.507391)	0.0	London city center
2	517	Ford Road, Old Ford	POINT(-0.033085 51.532513)	7.118	The 27 empty stations
3	852	Coomer Place, West Kensington	POINT(-0.202038700000003 5...	5.788	The 27 empty stations
4	846	Burgess Park Albany Road, Wal...	POINT(-0.0942844 51.48224)	3.629	The 27 empty stations
5	507	Clarkson Street, Bethnal Green	POINT(-0.059091 51.528692)	5.305	The 27 empty stations
6	554	Aberfeldy Street, Poplar	POINT(-0.005659 51.513548)	8.474	The 27 empty stations
7	850	Brandon Street, Walworth	POINT(-0.0915489 51.489102)	3.225	The 27 empty stations
8	851	The Blue, Bermondsey	POINT(-0.0625130999999897 ...	4.817	The 27 empty stations
9	523	Langdon Park, Poplar	POINT(-0.013475 51.51549)	7.956	The 27 empty stations
10	752	London Street, Paddington	POINT(-0.17371276 51.515117)	3.298	The 27 empty stations
11	21	Lansdowne Drive, Hackney Cen...	POINT(-0.062806212 51.53983...	5.759	The 27 empty stations
12	519	Teviot Street, Poplar	POINT(-0.011662 51.518811)	8.13	The 27 empty stations

Results per page: 50 1 – 50 of 55



#first hypothesis

Further in our pursuit, we are keen to identify the temporal nuances that drive customer engagement. Through a rigorous analysis of rental trends, we intend to pinpoint the best days and hours to maximize revenue. Armed with this insight, we can channel our resources towards peak demand periods, thus enhancing operational efficiency and delivering a seamless experience to our customers.



Row	start_dayofweek	Morning	Afternoon	Evening	Night
1	1	32.02	34.19	34.53	29.44
2	2	16.4	27.85	25.97	18.05
3	3	17.61	31.06	30.93	18.93
4	4	18.12	31.45	28.36	22.08
5	5	15.76	25.24	20.22	23.69
6	6	17.65	28.3	27.4	25.48
7	7	27.79	35.13	32.21	27.77

```
number of rides:
```

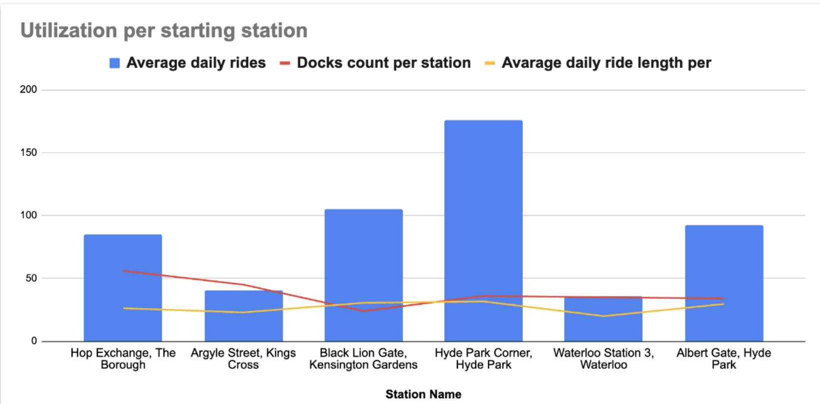
Row	start_dayofweek	Morning	Afternoon	Evening	Night
1	1	1230	6276	825	131
2	2	1419	2710	1127	96
3	3	1683	3936	1353	85
4	4	1686	2809	1156	98
5	5	1390	2055	660	45
6	6	1560	3491	820	104
7	7	1464	8107	1205	170

Additionally, customer preferences lie at the core of our investigation. By unearthing the best ride pairs (**which we haven't included in the conclusion page**), we hope to map the preferences of our riders and uncover popular routes. This knowledge is invaluable in designing optimal routes, adjusting station placements, and aligning our offerings with the unique demands of our clientele. In tandem with these questions, we are committed to addressing the issue of overcrowded stations, aiming to mitigate bottlenecks and ensure a smoother user experience.

Row	start_station_id	starting_name	AVG_ride_duration_p	AVG_daily_ride	dock_count	starting_geo_point
1	194	Hop Exchange, The Borough	26.24	84.59	56	POINT(-0.091773776 51.50462...
2	14	Argyle Street, Kings Cross	22.94	40.18	45	POINT(-0.123944399999999 5...
3	307	Black Lion Gate, Kensington Ga...	30.55	104.83	24	POINT(-0.187842717 51.50990...
4	191	Hyde Park Corner, Hyde Park	31.65	176.06	36	POINT(-0.153520935 51.50311...
5	154	Waterloo Station 3, Waterloo	19.98	35.67	35	POINT(-0.11282408 51.503791...
6	303	Albert Gate, Hyde Park	29.63	92.03	34	POINT(-0.158456089 51.50295...

Conclusion - best and worst starting stations

Our analysis reveal an interesting insight, as you can see at Hyde park and black lion gate station we have almost triple the rides each day from the count of the docking station, this means if we want to make better utilisation of these stations we will have to add more docking stations, where the average daily rides is low we can remove some docking stations to make the station utilization better.



As we steer our course through these hypotheses and questions, we envision a Zen City that is poised to set new industry benchmarks. The insights we derive from this exploration will empower us to refine station placements, enhance bike availability, and orchestrate a more enriching experience for our riders. As our strategies evolve based on these data-driven insights, we are well-prepared to navigate the dynamic urban mobility landscape and propel Zen City into a new era of excellence.

Chapter 4: Cleansing the Data Landscape for Informed Analysis

In this pivotal phase of our data analysis journey, we engaged in meticulous data cleaning and wrangling to ensure that our dataset was primed for insightful analysis. Guided by the principles outlined in the chapter's objectives, we embarked on an exploratory journey through SQL queries to rectify inconsistencies, mitigate outliers, and refine the dataset's integrity.

Our comprehensive data-cleaning journey commenced with the diligent identification of missing values, uncorrected data types, and null entries. By meticulously examining the dataset's attributes, we ensured that the foundation of our analysis was solid, eliminating the risk of skewed insights due to incomplete or erroneous data.

Our analysis extended further to address the potential outliers within the dataset. Employing rigorous statistical methods, we identified and carefully handled data points that fell outside the range of statistical normalcy. This approach allowed us to eliminate the influence of outlier data on our subsequent analysis, ensuring that our insights were truly reflective of the overall trends and patterns.

Through a series of strategically designed SQL queries within a CTE, we meticulously combined, cleaned, and aligned the data from the cycle_hire_new and cycle_stations_pro tables. This transformational process culminated in a refined dataset that was ready to be subjected to in-depth analysis, serving as the cornerstone of our quest to optimize Zen City's bike rental operations and elevate customer experiences. In the forthcoming chapters, we will harness the power of this curated dataset to derive impactful insights that will drive tangible enhancements in our bike rental service.
For further information about our process please consult our SQL script.

Chapter 5: Illuminating Insights through Data Analysis and Visualizations

In this pivotal chapter, armed with the refined dataset we meticulously curated, we delved into the realm of data analysis and visualization, unearthing a wealth of insights to propel Zen City's bike rental service. Our approach was multi-faceted, blending analytical rigor with strategic visualization to holistically understand user behaviors and the dynamics of the bike rental landscape.

Our journey commenced with univariate analysis, wherein we examined each variable independently to comprehend their individual distributions and characteristics. Utilizing histograms and box plots, we gained a comprehensive view of ride distances, durations, and station capacities. This provided a foundational understanding of the data's inherent trends, aiding in making informed decisions.

Building upon this foundation, we ventured into bivariate analysis, systematically exploring relationships between pairs of variables. Scatter plots enabled us to identify correlations between ride durations and distances, offering insight into potential user preferences. Meanwhile, bar charts visualized the relationship between different days of the week and ride counts, showcasing peak and off-peak periods.

To delve even deeper, multivariate analysis allowed us to untangle the interplay of multiple variables. Heatmaps exhibited the intricate relationship between ride distances, durations, and time of day, facilitating the identification of peak usage periods. Additionally, pie charts illuminated the distribution of ride types, shedding light on customer preferences for casual and subscribed rides.

Throughout our analysis, we incorporated a strong business sense, aligning each visualization with the overarching goal of optimizing bike rentals. By weaving domain knowledge into our interpretations, we contextualized our findings, facilitating strategic decisions to refine station placements, improve bike availability, and elevate the user experience. This chapter's synergy of data analysis and visualization was instrumental in enabling Zen City to navigate the urban mobility landscape with informed confidence, paving the way for innovation in the bike-sharing industry.

Finally, using our analysis we have reached **recommendations** for Zen:
Our mission was to find better ways to utilize our current existing bike grid, to analyze rental patterns and identify underutilized & overcrowded stations.
We found why there are empty stations, and how to better utilize the 6 available starting stations and how should we diversify our daily promotion to better drive traffic throughout the day.

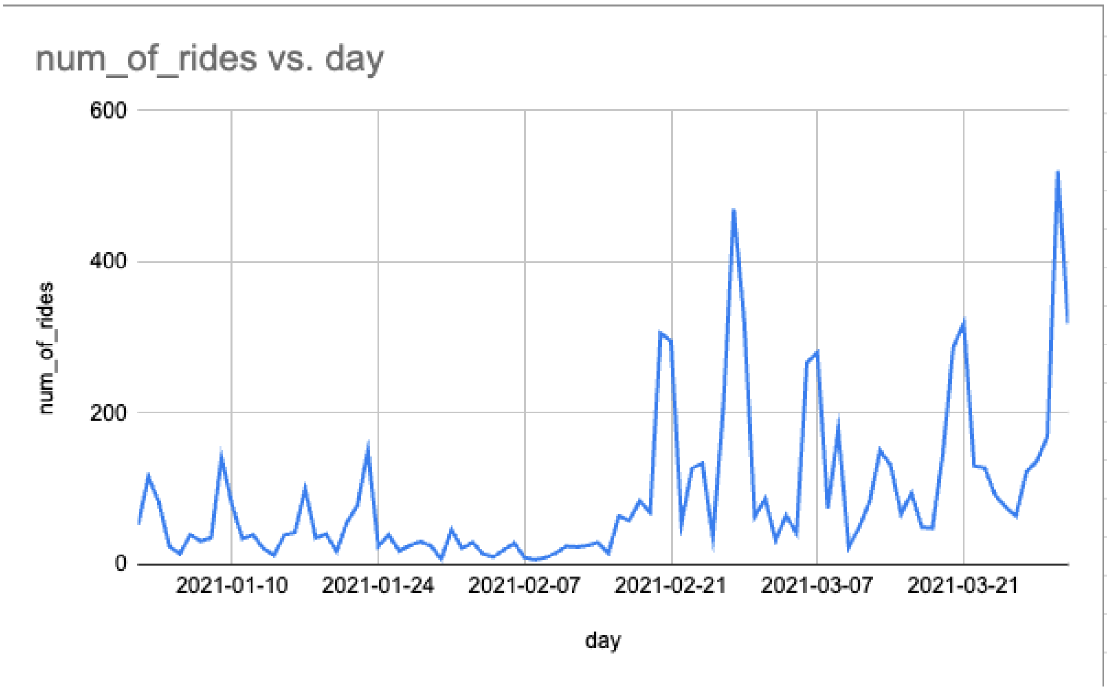
- Most of the rentals are for commuting to workplaces and colleges around the bustling core of London, Zen should focus on launching more stations closer to the city center to reach out to their main customers.
- The number of bikes available for rental and the number of empty docks in the station shows an imbalance, there is a higher demand than supply in most cases, Zen should focus on adding more available docking spaces to the 6 starting stations or open nearby starting stations to not miss out on any customers because of overutilization.

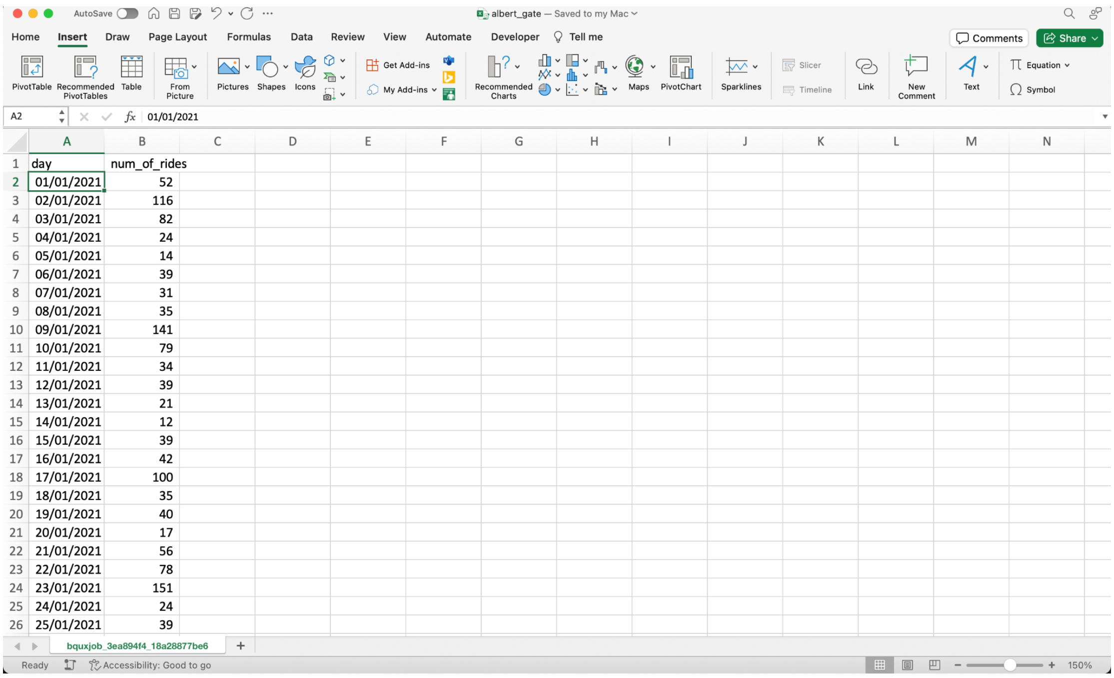
- While planning for extra bikes to stations the peak rental hours must be considered, i.e. 1–6 pm, in - addition, maintenance activities for bikes should be done at night due to low usage.
- offer dynamic pricing and promotions depending upon the day and hour to promote bike usage during the work week with a focus on the morning and night times.

Chapter 6: Envisioning the Future through Predictive Insights

In the final chapter of our data-driven journey, we harnessed the culmination of our statistical knowledge to answer a pressing prediction question: How many rentals would be made in April 2021 at the "Albert Gate, Hyde Park" (ID: 303) bike station? Drawing on the dataset's historical patterns and utilizing our analytical tools, we employed time-series analysis to forecast future demand with precision.
We've chose to use gretl (a statistical package) in order to run a OLS (Ordinary least squares) - a type of linear least squares method for choosing the unknown parameters in a linear regression model, because we really only have data regarding the number of rides per each day:
The equation: num_of_rides = $\alpha + \beta1 * time + \beta2 * time ^ 2$
And after uploading the data to gretl we were able to predict - forecast the daily number of rides for each day during April 2021, summing those values we will answer the question, that there will be 8,836 rides during April 2021 for "Albert Gate, Hyde Park" bike station.

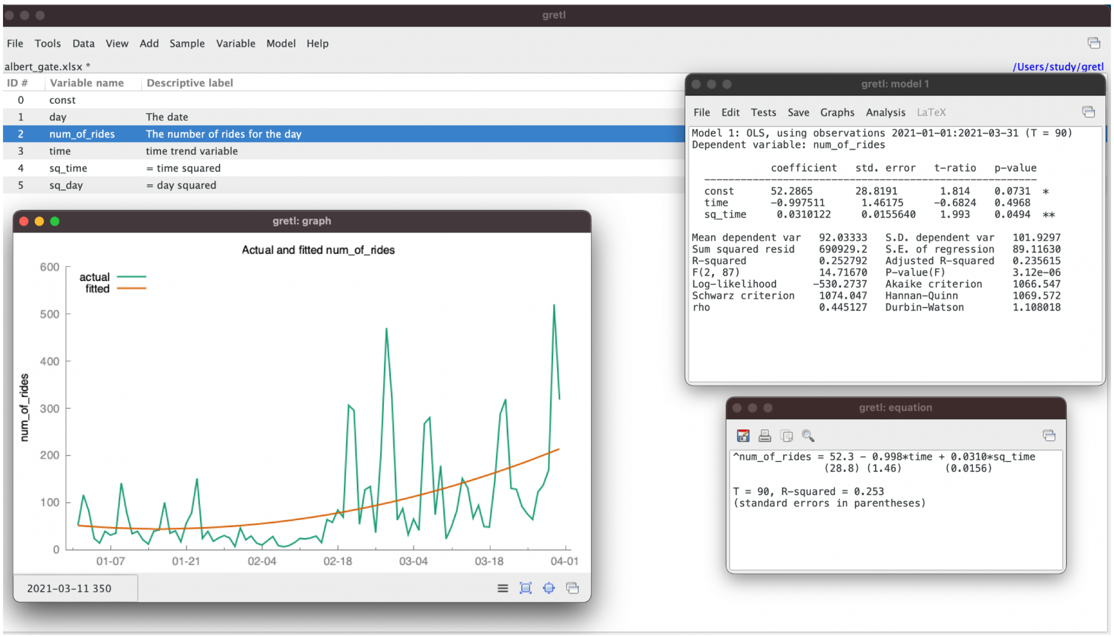
Row	day	num_of_rides
1	2021-01-01	52
2	2021-01-02	116
3	2021-01-03	82
4	2021-01-04	24
5	2021-01-05	14
6	2021-01-06	39
7	2021-01-07	31
8	2021-01-08	35
9	2021-01-09	141
10	2021-01-10	79
11	2021-01-11	34
12	2021-01-12	39



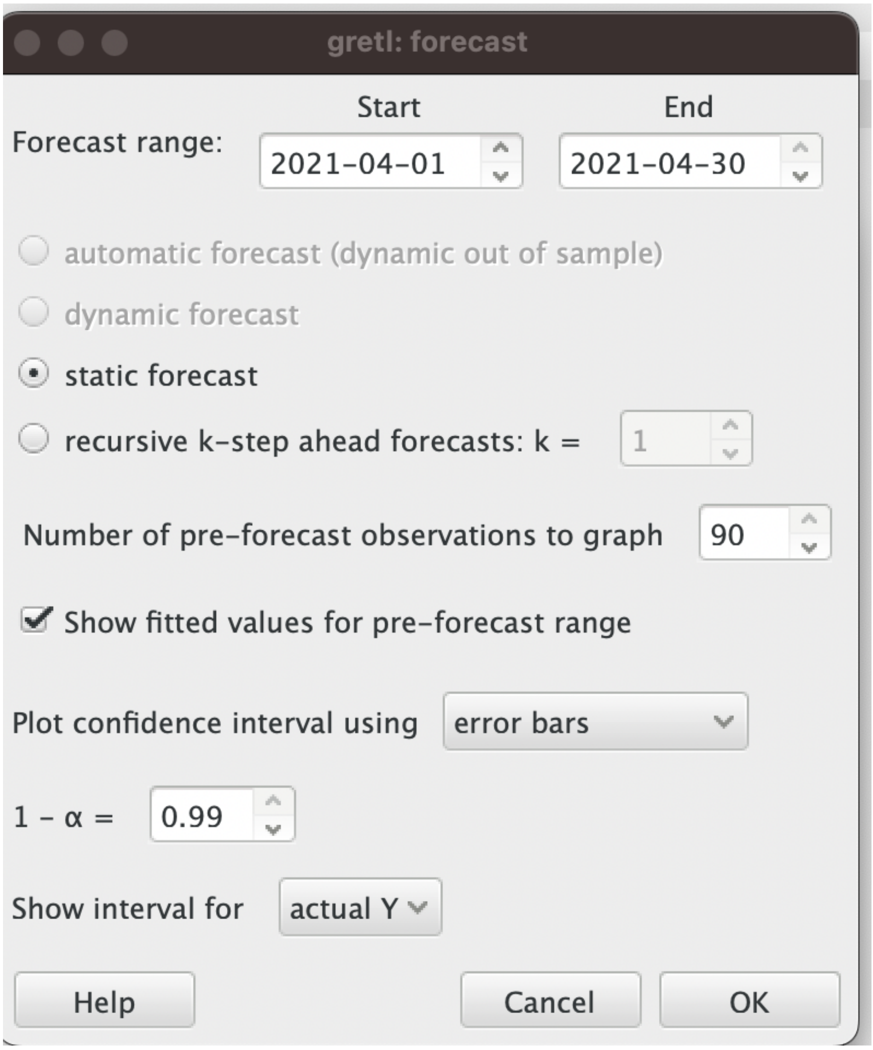


day	num_of_rides
01/01/2021	52
02/01/2021	116
03/01/2021	82
04/01/2021	24
05/01/2021	14
06/01/2021	39
07/01/2021	31
08/01/2021	35
09/01/2021	141
10/01/2021	79
11/01/2021	34
12/01/2021	39
13/01/2021	21
14/01/2021	12
15/01/2021	39
16/01/2021	42
17/01/2021	100
18/01/2021	35
19/01/2021	40
20/01/2021	17
21/01/2021	56
22/01/2021	78
23/01/2021	151
24/01/2021	24
25/01/2021	39

The equation: num_of_rides = α + β1 * time + β2 * time ^ 2

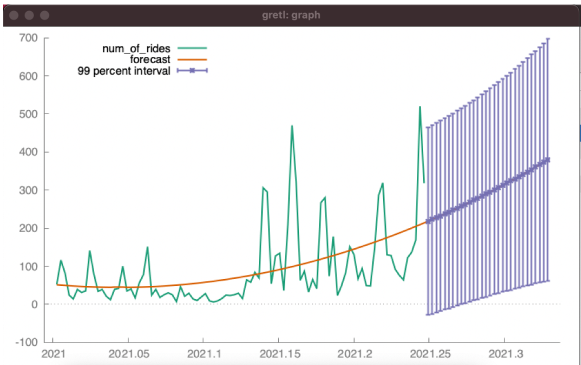


now, we use of model to predict - forecast the next 30 days: the month of April 2020:



(with a significance level of α = 0.01 -> 1-α = 99%)

gretl: graph



gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 2021-01-01:2021-03-31 (T = 90)
Dependent variable: num_of_rides

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const      52.2865       28.8191       1.814    0.0731   *
  time       -0.997511      1.46175     -0.6824   0.4968
  sq_time     0.0310122     0.0155640    1.993    0.0494   **

Mean dependent var   92.03333   S.D. dependent var   101.9297
Sum squared resid    690929.2   S.E. of regression   89.11630
R-squared            0.252792   Adjusted R-squared   0.235615
F(2, 87)             14.71670   P-value(F)           3.12e-06
Log-likelihood      -530.2737   Akaike criterion     1066.547
Schwarz criterion    1074.047   Hannan-Quinn         1069.572
rho                  0.445127   Durbin-Watson        1.108018
```

s): Full range 2021-01-01 –

gretl: forecasts

```
2021-03-10      23.00     131.11
2021-03-11      49.00     134.42
2021-03-12      83.00     137.80
2021-03-13     151.00     141.23
2021-03-14     131.00     144.73
2021-03-15      67.00     148.29
2021-03-16      94.00     151.92
2021-03-17      49.00     155.60
2021-03-18      48.00     159.35
2021-03-19     146.00     163.16
2021-03-20     287.00     167.03
2021-03-21     319.00     170.96
2021-03-22     130.00     174.96
2021-03-23     128.00     179.02
2021-03-24      92.00     183.14
2021-03-25      76.00     187.32
2021-03-26      64.00     191.56
2021-03-27     122.00     195.87
2021-03-28     137.00     200.23
2021-03-29     169.00     204.66
2021-03-30     520.00     209.16
2021-03-31     318.00     213.71
2021-04-01                218.32      93.660     -28.33 -   464.98
2021-04-02                223.00      94.067     -24.72 -   470.73
2021-04-03                227.74      94.502     -21.13 -   476.62
2021-04-04                232.54      94.969     -17.56 -   482.65
2021-04-05                237.41      95.466     -14.01 -   488.82
2021-04-06                242.33      95.996     -10.48 -   495.14
2021-04-07                247.32      96.560      -6.97 -   501.61
2021-04-08                252.37      97.158      -3.50 -   508.24
2021-04-09                257.48      97.791      -0.05 -   515.02
2021-04-10                262.66      98.461       3.36 -   521.96
2021-04-11                267.89      99.167       6.73 -   529.05
2021-04-12                273.19      99.912      10.07 -   536.31
2021-04-13                278.55     100.695      13.37 -   543.73
2021-04-14                283.97     101.517      16.62 -   551.32
2021-04-15                289.46     102.380      19.84 -   559.08
2021-04-16                295.00     103.283      23.00 -   567.00
2021-04-17                300.61     104.228      26.12 -   575.10
2021-04-18                306.28     105.214      29.20 -   583.37
2021-04-19                312.01     106.243      32.22 -   591.81
2021-04-20                317.81     107.315      35.19 -   600.42
2021-04-21                323.66     108.430      38.11 -   609.22
2021-04-22                329.58     109.589      40.98 -   618.19
2021-04-23                335.56     110.792      43.79 -   627.34
2021-04-24                341.60     112.040      46.54 -   636.66
2021-04-25                347.71     113.332      49.25 -   646.17
2021-04-26                353.87     114.669      51.89 -   655.86
2021-04-27                360.10     116.052      54.48 -   665.73
2021-04-28                366.39     117.479      57.01 -   675.78
2021-04-29                372.75     118.953      59.48 -   686.01
2021-04-30                379.16     120.471      61.90 -   696.42
```

	A	B	C	D
1	day	num_of_rides		
102	11/04/2021	268		
103	12/04/2021	273		
104	13/04/2021	279		
105	14/04/2021	284		
106	15/04/2021	289		
107	16/04/2021	295		
108	17/04/2021	301		
109	18/04/2021	306		
110	19/04/2021	312		
111	20/04/2021	318		
112	21/04/2021	324		
113	22/04/2021	330		
114	23/04/2021	336		
115	24/04/2021	342		
116	25/04/2021	348		
117	26/04/2021	354		
118	27/04/2021	360		
119	28/04/2021	366		
120	29/04/2021	373		
121	30/04/2021	379		
122	**The sum for April 2021:**	**8836**	start_month	num_of_rides
123			1	1491
124			2	2502
125			3	4290
126			4	**8836**
127				
128				
129