BIKE SHARE TREND ANALYSIS

Submitted by: Bhakti Raichura, Khushee Thakker, Shreya Srirama, Varsha Srinivasan

Abstract: It is estimated that there are more than 2 billion bikes used in the world and this number is constantly growing. It is also known that over 50 percent of the world's population knows how to use a bike. Owing to social and health benefits, people are now moving from car-sharing to the bike-sharing model and even further, e-bikes. The e-bike market is estimated to be valued at over $20B and continues to grow by the year 2030. In such a world where there are several bike and e-bike riders, our focus is to analyze the users' usage of bikes, their dependability on the weather conditions, and correlations with other nodes such as stations, trips, users' age groups. The data analysis is done in MySQL and Python, AWS for data warehousing and ETL and finally data visualization is done in Tableau. Through our trend analysis, we were able to gain insights into e-bike ride patterns and deduce suggestions to increase profitability for any e-bike-sharing company like Lime, Lyft etc.

Key Terms: Data warehouse, Data models, ETL process, Data Normalization, Visualization, Database, Datalakes

Motivation

In the world of increasing global temperature, bike riding not only provides health benefits but shifting trips from cars to bikes also helps reduce congestion, air pollution, and CO2 emissions. There are immense benefits of making this transition from driving cars to riding bikes. Through this project, we aim to highlight some of those benefits and highlight the patterns that can boost up the sales of such bikes. Our motive is to encourage not only the youth, but everyone to switch to riding bikes and e-bikes. Also, post 2020, the world saw an immense shortage of chips used in major vehicles creating a downfall for the automobile industry making the common man switch to riding bikes. We take this opportunity as helping the companies and manufacturers of such bike and e-bikes by identifying trends in the ride sharing of a typical user.

There are several studies done on bike sharing systems and predicting the usage of how external factors influence the patterns of e-bike ride sharing in people.

Literature Survey

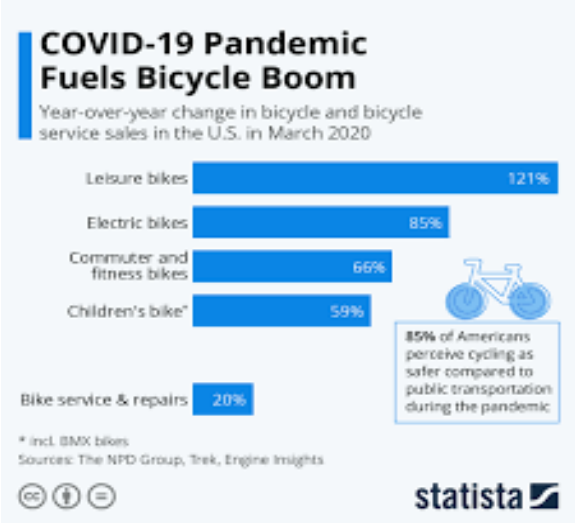
Nankervis's study shows the weather effects (long and short terms) on bicycle commute for tertiary students in Australia and Melbourne. The dataset focuses on students who are young and healthy, who commute using bicycles to their universities. The dataset includes various facets like places the students commuted, distance covered, riding time, % of riders based on months and semester weeks, etc. and helps identify the trends in seasonal behavior.

LarsBöcker, MartinDijst and JanFaber's research helps identify effect of weather conditions affecting choices of transport mode, travel experiences along outdoor thermal perceptions. Insights of users' emotional experiences influence travel behavior in terms of public and active mode transports are highlighted in this study

Elliot Fishmana, Simon Washington, Narelle Haworth, Angela Watsonc's study on the factors affecting membership in Australia's bike-sharing plan shows findings in the prospective increase of potential users in the future and the reliability of the transport system using a logistic regression model. It draws conclusion on several external and internal factors which decide whether or not a user will purchase membership to a bike share model.

Industry influence

The current automobile industry is at an all-time cost high. This is owing to the problems in the supply chain process post Covid-19. Due to this, people have switched gears to riding bikes and e-bikes and the below chart shows trends in purchase patterns of bikes across the United States. Studies have shown that people felt more comfortable riding bikes during the pandemic than using public transport



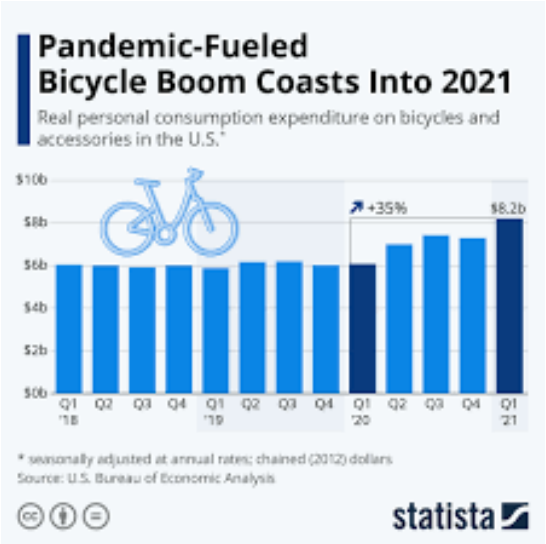


Fig 1

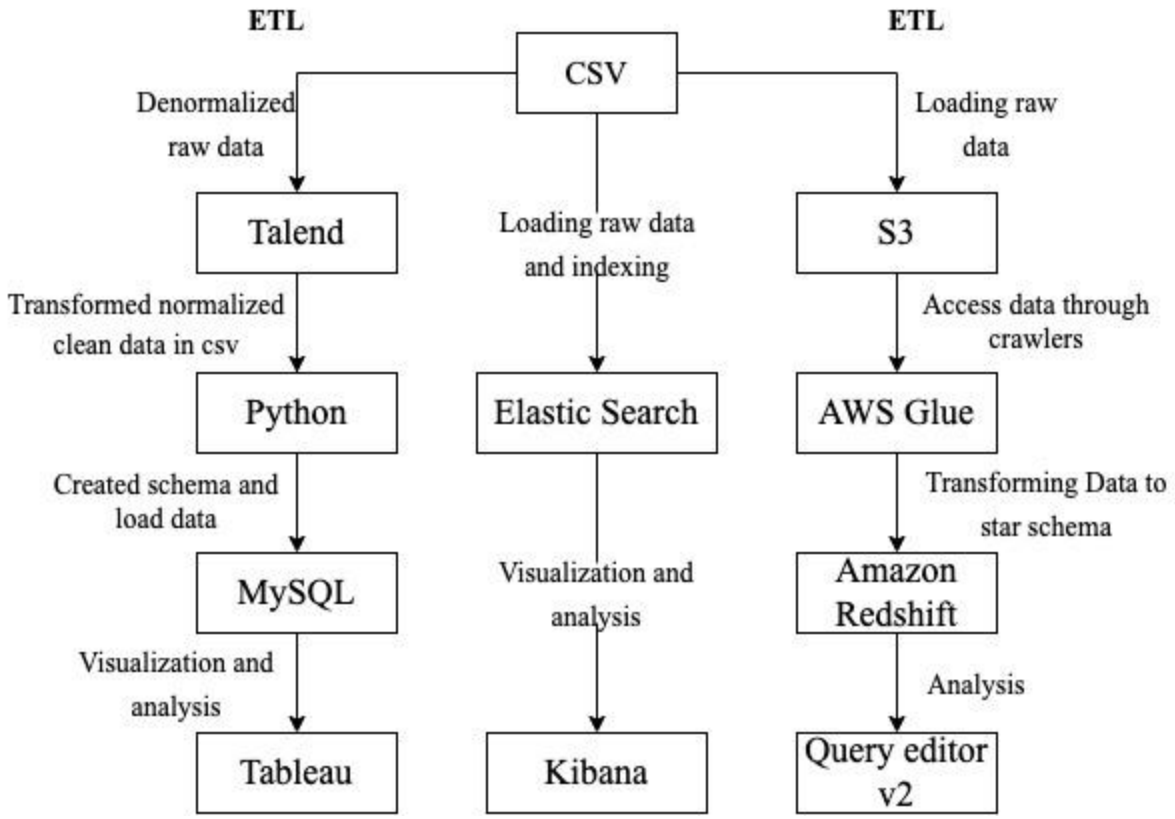
Fig 2

Fig 1. shows how the sales of bikes have boomed during Covid-19. It also depicts how electric bike sales have increased in the US starting March 2020. Another depiction in Fig 2. shows quarter over quarter growth in bike riding in 2021.

Several statistics and surveys show a booming demand for e-bikes in the coming year, estimatingET the industry to be over $60B by the year 2030.

Project Flow and design

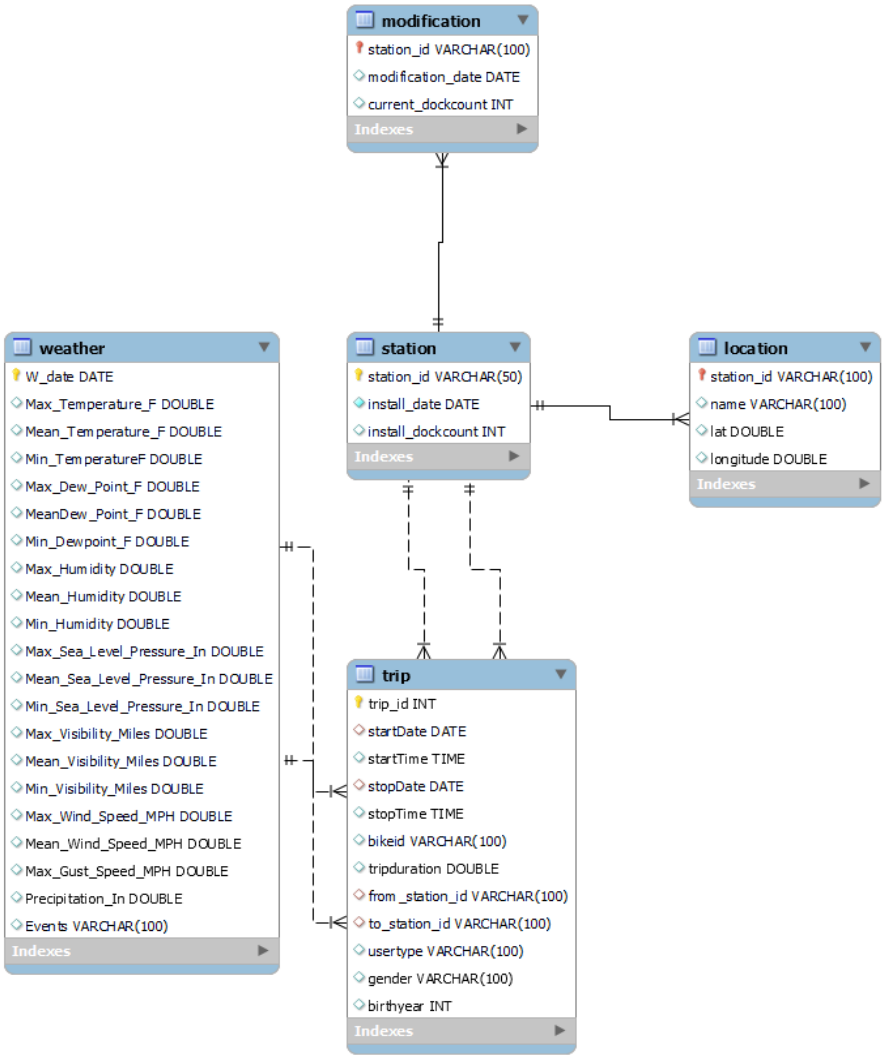
We used multiple tools and technologies in the project. The multiple csv files were loaded to python through the Mysql connector. We used several operations in Mysql to perform analysis. The queries used in MySQL and python analysis made use of operations like JOIN statements, UNION, Window functions, Aggregates etc. Next, we used AWS as our data warehouse and S3 datalake for storage and Glue for ETL. Finally, the dashboarding and visualization is done in Tableau. Elastic search as an additional NoSQL database was also used in the project.



Data Model

a) ER diagram

The below Entity- Relation diagram represents 5 tables used in the project and their attributes and relationships. The tables used are trip, station, weather, modification and location



Datawarehouse: