Los Angeles Metro Bike Share Visualization

Chengxiang Duan[1], Jieqiong Pang[2], Linle Jiang[3], and Shengjia Wu[4]

[1] duanc@usc.edu
[2] jieqiong@usc.edu
[3] linlejia@usc.edu
[4] wushengj@usc.edu

Abstract. LA Metro bike share is a bicycle sharing system. With the increasing use of bike sharing system by the LA residents, service providers need to constantly optimize their operations, including station planning and bike deployment, to meet the customer needs and reach profit growths. However, existing visualization works (i.e., dashboards) did not provide a comprehensive service management tool to support service providers. In this project, we targeted service providers as our end users and built a service monitoring tool, aiming at providing actionable insights for our users to help them make informed decisions about station planning and bike deployment.

Keywords: Visualization · LA Metro Bike Share · Dashboard · Operation Analysis

1 Introduction

Metro Bike Share is a bicycle sharing system in the Los Angeles, California metropolitan area, launched on July 7, 2016. As a profitable business, along with the increasing use of metro bike services, administrators and operators need to constantly optimize their operations, including station planning and bike deployment. To make informed decisions, users not only need to understand the operation situation, but also evaluate how prior decisions changed the situations in the subsequent quarter. Data Visualization on the Metro bike share data can help users understand the data and extract information effectively. Here we built a service management tool, generating actionable insights for our target users to help them achieve their key business objectives, especially in logistic optimization. Given that existing work didn't visualize time-series or geo-spatial data simultaneously in the same dashboard, our work aims to fill this gap and build a comprehensive visualization tool to meet our users' needs.

2 Related Work

Prior work visualized the most popular destinations using mapbox [4] (Capital Bikeshare Destinations by Station), showed descriptive statistics [5] using simple charts (NYC Citi Bike Dashboard), or designed activity site heatmap [2] dashboards to indicate station popularity (BIXI Station Density Map Comparison). However, most of these works only provided descriptive statistics at a time snapshot, and it is impossible to predict or evaluate decision outcomes without time series data.

We intend to fill this gap in our project. The biggest advantage of our project is to have a clear user profile. Unlike previous work, which did not allow users to make or evaluate decisions, we developed a service monitoring tool, specifically tailed to the needs of service providers, offering actionable insights for service planning. This is made possible by including multiple years of data for visualizing trends and changes. Moreover, we visualize the data over map thanks to the availability of geo-spatial attributes, further supporting our users to extract patterns from the visualizations, make evaluations over prior decisions, and maximize future decision outcomes.

3 Data

The project included three datasets: LA Metro Bike Share Trip dataset, Station dataset, and LA County City Boundaries Geo dataset. The trip and station datasets were collected from the the official Bike Share Website [1]. Each trip dataset consists of all bike rental record in a quarter, containing trip duration, start/end station, trip date attributes. Overall, 16 trip datasets from 2016Q3 to 2020Q2 yield 1.06M rows of records. The station dataset provides information about 292 bike stations in LA, including station coordinates, station id, go-live date, etc. The LA County City Boundaries Geo dataset was downloaded from the LA GeoHub Website [3], in Topojson format, which was used to draw the d3 vector map.

We cleaned the trip and station dataset by unifying the attribute names and units, while also deleting unrelated properties, such as "plan duration." We also deleted all information about the virtual stations, which was operated only for bike arrangement purposes. We then extracted information from trip datasets, such as, in a quarter, the sum of trip duration between two stations, the popularity of a station. As a result, station related statistics were combined with the station dataset, generating station Geojson files for each quarter. Then, we sort the station by popularity and record the top ten and the bottom ten stations each quarter. Additionally, a separated Geojson file summarizing station popularity of all quarters was generated for heatmap visualization.

4 Approach

Our design process mimicked actual web app development process. We first identified bike share service providers as our target end users. Then we researched and analyzed our users, trying to understand what they need by brainstorming and establishing the user cases. This process defined the scope and mission of this project. And we ideated features and solutions to address these user needs. Next, we applied wireframing and prototyping techniques to design a prototype for this website using Adobe XD, a user experience design tool for app designs. And we structured the layouts and styles for our website based on the design. Finally, we implemented the visualizations and inserted them into the predetermined positions with specific features.

The visualizations are at the region- and station-level. In particular, by looking at region-level visualizations, users are able to evaluate prior decisions on the stations at a high level, including but not limited to, whether the opening of new stations drew more usages of bikes, relieved the intensive use of nearby stations, and whether such measures yielded positive outcomes. In this visualization, an overview map (Fig. 1). showing all stations is provided. Users click on each region could understand how station number, total and average trip frequency per region changed overtime with the help of a line chart (Fig. 1). This helps users evaluate their prior decisions and narrow down their analyses to a particular time range and region. To further limit the scope of the analysis, we also provide an interactive heatmap (Fig. 1). Once users select a desired region, the heatmap automatically centers and zooms to the specific region. Users could then drag the slider to select the specific quarter, and review any changes occurred on station setup and station popularity overtime, providing additional insights for decisions evaluation at particular area and time, and further pinpoint the area that requires attention. In this case, region-level visualizations serve as a starting point, helping users narrow down to specific areas to attend to, and focus subsequent analyses on these areas. Yet, it is the station-level dashboard that provides details on the performance of each station and actionable bike deployment insights. Users could specify a quarter they want to focus on this page. At the top left, users have two bar charts showing the most and least popular stations (Fig. 2), providing additional station candidates they could attend to. Users could choose one station and obtain comprehensive statistics on how the station was used, as well as the bike use status. Specifically, users receive a vivid map visualization indicating all other stations connected to the station of interest and the weights of these connections. To further support such between-station analysis, we also provide an interactive donut chart (Fig. 2) to visualize the top connected stations, with respect to in- and out-station relations. In terms of within-station analysis, which supports bike deployment decisions, we not only include dashboard cards reflecting in- and out-station bike counts and trip duration measures, but also provide a bar chart for trip duration distribution (Fig. 2) and a line chart (Fig. 2) for busy time distribution. All of these combined provide actionable insights on how users could optimize their bike deployment.

In terms of the visual design, we seek depth, clarification and boom effect of the dashboard. First, each page has a grid layout to be responsive. We also simplified the information in dots, lines, bars, arcs, and summarized annotated big numbers. Furthermore, we used safe fonts and colors consistently. Font family was set to increase the availability for users to use in the browser. Finally, to achieve the 5-second rule, we divided the two main topics into two separated tags, and used line charts and maps to show key trends.

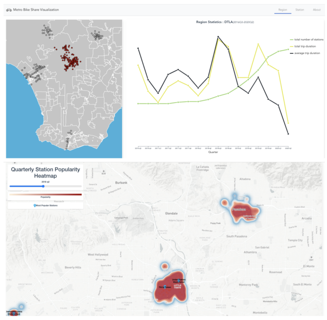

Fig. 1. Region Level Analysis Page.

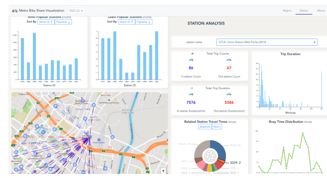

Fig. 2. Station Level Analysis Page.

5 System

In the project, we used Vue.js framework [6]and arranged the folder structure to separately store Vue components, Javascript files, SCSS style files, and data files, making them in a hierarchy structure for a collaboration purpose. Moreover, we kept two branches, "main" and "dev," in Github. The "main" branch was mainly for publishing stable versions, while the "dev" branch was for development increment purposes. In the local Git store, we work on separate feature branches and push the new features to the "dev" branch. This workflow also helped us collaborate efficiently. In our website's "station" page, the "selected station" had to be set and tracked by many Vue components, so we used Vuex [7]to track this common variable. Users could choose any station on the page in various ways, such as using the dropdown menu, in the map, in the bar chart, and the donut chart. Per selection, new data would be loaded, flowed to corresponding Vue components, and triggered the chart generator.

We used flexbox, grid, and sass to achieve a responsive effect. We set breakpoints in css for large, medium, and small window sizes and arranged the grids differently for each window size. As a result, when users change the window size, the components move to fit the window cascadingly. We also took advantage of the features in sass. We used mixins to apply snippets of style to multiple elements, which was neat and flexible. We also set variables for color schemes. When defining styles, we could simply claim the variable instead of color in html color model (eg: #ffffff).

We used simple and/or advanced layouts to build responsive and interactive D3 diagrams with animation transitions, and generated generic template JavaScript files for the line, bar, and donut diagrams. These template files are reusable simply by passing in different data parameters.

In the region page, we used d3 to construct a LA county map that visualizes the geo-locations of stations. An interactive effect was included to enable region selection. We also drew an interactive heatmap to visualize quarterly station popularity data using Mapbox, which also listens to the users' selection in the d3 map, and automatically centers and zooms to the specific region, accordingly. The interactive effect allows users to use a slider to navigate through different quarters of data. Specifically, all layers for the quarters were created simultaneously after the data was loaded, with the visibility set to 'none' as the default. Each quarter has three layers, including a marker layer (i.e., most popular stations), a heatmap layer, and a circle layer. The size of the layers adjusts automatically based on the zoom level. As users drag the slider, the corresponding quarter layers are set to 'visible' to show the contents, with the others in 'none'. In the station page, we used Leaflet to create a station distribution map. An interactive effect was also included to enable users to focus on individual stations. And we specified the css details to create lines with animated effect showing the in- and out-station traffics in relation to the station of interest, with the line width indicating different frequencies.

6 Conclusion

To the best of our knowledge, this is the first visualization project specifically designed for bike share service providers, with the features and designs aiming at providing actionable insights for our users and to help them make informed decisions about station planning and bike deployment. In this project, we mimicked real-world app development processes, incorporating web design processes, such as identifying target users, ideation, wireframing and prototyping to determine the visual and functional aspect of our artifact; and following a mature software development branch workflow (i.e., Git). We also limited the use of Vue to layout control and data streaming tool, while creating reusable chart generator codes, which are customizable by feeding in data and parameters, for the same types of charts (e.g., bar charts and line charts).

Hereafter, we want to automate the ETL process with a PostgreSQL database. Eventually, with a simple click, users could see the real-time visualization.

References

1. metro bike share. https://bikeshare.metro.net/about/data/, last accessed 29 Nov 2020
2. Bixi station density map comparison. https://public.tableau.com/profile/mergeyourdata#!/vizhome/BIXIStationDensityMap/BIXIStationAnalysis, last accessed 29 Nov 2020
3. La county city boundaries. https://geohub.lacity.org/datasets/lacounty::la-county-city-boundaries?geometry=-119.879%2C33.678%2C-116.410%2C34.474, last accessed 29 Nov 2020
4. Capital bikeshare destinations by station. https://public.tableau.com/profile/karl3594#!/vizhome/CapitalBikeshareDestinationsbyStation/DestinationsbyStation, last accessed 29 Nov 2020
5. Nyc citi bike dashboard. https://public.tableau.com/profile/taishi.matsuda#!/vizhome/BikeSharing_15909395256870/NYCCitiBikeAnalysis, last accessed 29 Nov 2020
6. vuejs. https://vuejs.org/, last accessed 29 Nov 2020
7. vuex. https://vuex.vuejs.org/, last accessed 29 Nov 2020