

Fall 2021 Master of Statistical Practice - Joint Project

36-601 Perspectives in Data Science
36-611 Professional Skills for Statisticians

Predicting bike share availability

October 25th, 2021

Overview

Bike-sharing platforms are on the rise, providing affordable alternatives to popular existing forms of transportation for short trips. Furthermore, bike-sharing businesses address community issues including traffic congestion, and serve the public through environmental, health, economic, and social benefits. Our client is a third-generation bike-share technology company, providing scaled services in limited major metropolitan markets. Our client has continually expanded into new, lower-demand locations and offered a variety of membership options to increase accessibility and broaden its customer base.

Rising costs of independent contractor labor, costs for exclusive use of public state & local territories, technical and hardware issues at solar-powered docking stations, and user experience/reliability problems of the core platform have all chipped away at our client's growth strategy and profitability. The company is evaluating strategic options to generate alternate sources of revenue, while seeking short-term solutions to improve the user experience and sustain existing market share for the core business.

Description of the Client Problem:

Our Client has engaged CMU's MSP program to address a significant pain point in the user experience: ensuring bike availability across their Washington, DC locations. Our client has years of existing usage data, and would like to leverage that data to lower operational costs associated with the unpredictable nature of bike availability.

Project Scope

1. **Build a model** to predict availability of bikes by station at any given future date/time, based on historical information available. The model is intended to improve bike availability by scheduling contractors who reshuffle bikes from/to bike stations.
2. **Create a presentation** for your client to understand the new model – this should address, at minimum:
 - a. An overview of the dataset used
 - b. A non-technical summary of the methodology used to create a prediction model
 - c. A summary of the accuracy of the prediction model
 - d. A plan to maintain the model going forward
 - e. Actionable recommendations describing how the model can be used to lower operational costs or increase revenue
 - f. Next steps for a phase 2 project
 - g. A technical appendix that explains the statistical details of the methodology used in the prediction model

Specifications

Historical data on Capital Bikeshare usage can be accessed at this URL:

<https://www.capitalbikeshare.com/system-data>

Additional datasets are encouraged to improve the accuracy of the prediction model.

Project Milestones

1. Class Q&A - **36601 - Monday, 10/25/2021, 10:10am** (in-class activity)

Questions related to this project will be addressed during Perspectives in Data Science class lecture.

2. Project Delivery Leading Practices

This will count as 20% of your HW#7 grade for 36611.

Submit link to Initial Project Preparation deliverables (one document, multiple tabs) by 5:00pm via email to jspm@cmu.edu

2.1.1 Project Charter - October 29th, 2021

2.1.2 Project Plan - October 29th, 2021

2.1.3 Requirements - October 29th, 2021

Submit Status Report links on due date and update Project Plan with actuals, to be submitted by 5:00pm via email to jspm@cmu.edu

2.2.1 Status Report 1 - November 4th, 2021

2.2.2 Status Report 2 - November 12th, 2021

2.2.3 Status Report 3 - November 19th, 2021

3. Class Brainstorm - **36601 - Wednesday, 11/3/2021, 10:10am** (in-class activity)

As a class, brainstorm together related to questions that arise during the EDA process, potential machine learning methodologies to try, potential explanatory variables, and potential risks to avoid.

4. EDA Presentation (part of **36601 Final project**) - **EOD 11/12/2021**

This will count as 25% of the final project grade for 36601.

5- 10 slide EDA presentation due to 36601 via Gradescope; 5 minute presentation to present 1 slide from presentation document, in class on 11/15/2021. EDA should include summary statistics and visualizations describing bike availability and any other relevant features.

5. Final Presentation & Models (**36601 Final project**) - **EOD 11/30/2021**

This will count as 75% of the final project grade for 36601.

Final presentation and code addressing Project Scope due to 36601 via Gradescope.

Presentations must be accessible via Project Repository for jspm@cmu.edu no later than November 30th.

6. Oral Presentations - ***Scheduled in 601 and 611 December 1, December 3***

Each team will get 15 minutes to present their project to the class - project presentations will occur during 36601 and 36611 classes the week beginning November 29, 2021.

Final presentation (written and oral), focused on delivery, professionalism, clarity and communication to a non-technical audience, will count as 80% of your HW#7 grade for 36611.