First Capstone Project Proposal

Project Idea 1: Maximizing fleet efficiency for Capital Bikeshare System

PROPOSAL DETAILS

1. **What is the problem you want to solve?**

I plan to analyze Bike Sharing system program data hosted by UCI

http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset

This dataset contains the hourly count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information. The dataset contains nearly 17K records with 16 attributes.

instant	dteday	season	yr	mnth	hr	holiday	weekday	workingda	weathersi	temp	atemp	hum	windspee	casual	registered	cnt
1	1/1/11	1	0	1	0	0	6	0	1	0.24	0.2879	0.81	0	3	13	16
2	1/1/11	1	0	1	1	0	6	0	1	0.22	0.2727	0.8	0	8	32	40
3	1/1/11	1	0	1	2	0	6	0	1	0.22	0.2727	0.8	0	5	27	32
4	1/1/11	1	0	1	3	0	6	0	1	0.24	0.2879	0.75	0	3	10	13
5	1/1/11	1	0	1	4	0	6	0	1	0.24	0.2879	0.75	0	0	1	1
6	1/1/11	1	0	1	5	0	6	0	2	0.24	0.2576	0.75	0.0896	0	1	1
7	1/1/11	1	0	1	6	0	6	0	1	0.22	0.2727	0.8	0	2	0	2

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. Currently, there are over 500 bike-sharing programs around the world.

Objective is to improve management & efficiency of Capital bicycle sharing business by:

a)   Summarize spatial movement of fleet

b)   Define high and low demand times

c)   Improve fleet management (e.g. placement of bikes or planning maintenance) based on spatial demand & usage

d)   Predict user type (registered vs. casual) based on time of day or week

2. **Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?**

The client is Capital Bikeshare System and this research is timely, as Capital Bikeshare System is in the process of accepting bids from for-profit bike share programs in order to phase out their non-profit operation. Considering the current circumstances of the program, it is beneficial to analyze bike share user behavior to determine if certain business decisions might increase operational efficiency or increase revenue potential.

Questions for the analysis:

How does weather - temperature, precipitation, humidity, and wind speed - affect bike use?

How does bike use vary between weekdays and weekends?

How does use behavior differ between members and nonmembers? Behavioral use questions to include mean, median, mode, and range for ride length, day of week, time of day, location of departure, and location of arrival.

Potential applications of the analysis:

Optimization model for maintaining bike stock at docking stations based on demand forecasting.

Predictive model for volume of bike use based on a combination of weather variables.

Exploration of shifting from flat rate pricing structure to dynamic pricing for revenue optimization.

Predictive model for best future locations of bike docking stations.

3.   **What data are you going to use for this? How will you acquire this data?**

Capital Bikeshare posts quarterly data reports of bike trip times, start and end locations, and type of user (registered or casual). Each trip is on one line of data. These data are readily and publicly available at https://www.capitalbikeshare.com/system-data. The data look like this:

Duration	Start date	End date	Start station	Start station	End station r	End station	Bike number	Member type
367	4/1/18 0:00	4/1/18 0:06	31103	16th & Harva	31214	17th & Corcc	W20968	Member
653	4/1/18 0:00	4/1/18 0:11	31201	15th & P St N	31503	Florida Ave &	W20887	Member
598	4/1/18 0:02	4/1/18 0:12	31268	12th & U St N	31251	12th & L St N	W00133	Member
2015	4/1/18 0:03	4/1/18 0:37	31505	Eckington Pl	31228	8th & H St N	W20534	Casual
563	4/1/18 0:04	4/1/18 0:14	31214	17th & Corcc	31223	Convention (	W23117	Member

4.   In brief, outline your approach to solving this problem (knowing that this might change later).

The first goal is to exhaustively determine the various relationships between bike use and weather variables.

1. Data wrangling will occur to clean datasets to applicable variables and check for inconsistencies in Capital Bikeshare dataset as some stations may have switched mid-season.
2. Exploratory Data Analysis (EDA) will occur to check for possible trends and/or correlations between bike usage characteristics and weather variables. The basic preliminary questions should be confidently answered after this stage.
3. Machine Learning, Statistical Modeling, and/or Algorithmic applications will be considered in hope of developing one of the possible applications mentioned above. As of this proposal writing, the most realistic application would be to develop daily bike use predictions based on a specified set of weather variables. It would be very appealing to include this application idea as a dashboard of some kind in the project deliverables.
4.

I intend to visualize the data using mapping techniques, and hope to develop some schematics to represent fleet movement. Histograms indicating high and low demand times at the various bike storage locations will help to develop more thorough analysis such as predicting user type based on time of use.

5.   What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

The final products will include code generalized for use with any quarterly report of bike usage data, as well as a slide presentation to describe the methods and benefits to Capital Bikeshare.