

Project Proposals

Proposal 1: - Citi Bike, NYC

Data set: <https://www.citibikenyc.com/system-data>

Proposal 2: - Yelp reviews, NYC

Data Set: <https://www.yelp.com/dataset/challenge>

Proposal 3: - Airbnb host reputation.

Data Set: <http://insideairbnb.com/get-the-data.html>

Project Proposal 1: - Citi Bike, NYC - Repricing case study and Twitter Sentiment analysis

Client - Citi bike NYC

Summary - Citi Bike is the largest bike share program in us, with 10,000 bikes and 600 stations across Manhattan, Brooklyn, Queens and Jersey City. It was designed for quick trips with convenience in mind, and it's a fun and affordable way to get around town. Everyone knows that bike sharing is the answer to many environmental and urban transportation issues, yet it's not mainstream in us. I am being asked by the senior executive team to use data science to recommend 3 key action item to increase the company business immediately.

Project Goals -

- User Sentiment analysis from twitter and identify most common customer issues.
- Identify which of the current plan can be repriced to get 5% increase on revenue with no customer impact.
- Propose a new monthly pass pricing for office commuters.

Approach -

- Business use case analysis and EDA
- Clustering and Geospatial analysis
- Regression and Machine learning model for prediction
- NLP -User sentiment analysis based on twitter hashtags and yelp reviews

Data -

- Bike trip data provided by Citi bike: <https://www.citibikenyc.com/system-data>
- Daily weather data from open sources

Deliverables -

Source code in python notebook on github repository
Project Summary report for client
Technical design document

Project Proposal 2: - Yelp reviews, NYC – User Sentiment analysis for a business in NYC

Client - Yelp Inc.

Summary - Yelp publishes crowd-sourced reviews about local businesses. The company also trains small businesses in how to respond to reviews, hosts social events for reviewers, and provides data about businesses, including health inspection scores. So the biggest challenge for yelp lies is what's in a review? Is it positive or negative? Yelp reviews contain a lot of metadata that can be mined and used to infer meaning, business attributes, and sentiment. I am being asked by the senior executive team to use data science to do a sentiment analysis of reviews for local business and identify overall satisfaction score and predict future sentiment for a business.

Project Goals -

- User Sentiment analysis from Yelp reviews and identify most common customer issues.
- Propose a new satisfaction score for each business based on review sentiment.
- Train a machine learning model to predict future sentiment for a business

Approach -

- Business use case analysis and EDA
- Clustering and Geospatial analysis
- Regression and Machine learning model for prediction
- NLP -User sentiment analysis based on user reviews

Data -

- Yelp review data: <https://www.yelp.com/dataset/challenge>

Deliverables -

Source code in python notebook on github repository
Project Summary report for client
Technical design document

Project Proposal 3: - Airbnb – Why do users like (or don't like) Airbnb experience and how likely they will book again.

Client - Airbnb Inc.

Summary -

In January 2018 the Airbnb had over 3,000,000 lodging listings in 65,000 cities and 191 countries. The company does not own any lodging that means they have to rely on the hosts to maintain the quality of overall user experience. Reputation management is a huge challenge for both host and Airbnb. I am being asked by the senior executive team to use data science to do a sentiment analysis of user reviews and identify an overall happiness score and using this eventually predict how likely a property going to be booked in near future.

Project Goals -

- User Sentiment analysis from user reviews and identify most common things user like or dislike about their overall experience.
- Train a machine learning model to predict how likely a property going to be booked in future.

Approach -

- Business use case analysis and EDA
- Clustering and Geospatial analysis
- Regression and Machine learning model for prediction
- NLP -User sentiment analysis based on user reviews

Data -

- Airbnb data: <http://insideairbnb.com/get-the-data.html>
- Daily weather data from open sources

Deliverables -

Source code in python notebook on github repository
Project Summary report for client
Technical design document
