

Presented By: Group 10

Delina Ivanova

Lynda Yearwood

Dickson Choy

Abby Sivan

Asaf Mendelovitz

BIXI: IPO Analytics

Table of Contents

[Introduction 3](#_Toc519450488)

[Analytics Requirements 3](#_Toc519450489)

[Data Preparation & Analysis 4](#_Toc519450490)

[Data Source 4](#_Toc519450491)

[Data Quality 4](#_Toc519450492)

[Data Cleaning Methods 4](#_Toc519450493)

[Tools & Programs Used 4](#_Toc519450494)

[Coding: Highlights 4](#_Toc519450495)

[Statistical Methods 5](#_Toc519450496)

[Data Visualization 5](#_Toc519450497)

[Analysis & Conclusions 5](#_Toc519450498)

Table of Appendices

[Appendix 1: TBD 6](#_Toc519450532)

Table of Figures

Figure 1: Code for removing trips with the same start and end stations 4

Figure 2: Method for Calculating Trip Distance 5

Table of Tables

Table 1: Ridership Raw Data 3

Table 2: Station Raw Date (Station Name Dropped) 3

Table 3: Master Dataframe (Sample) 4

# Introduction

BIXI Montreal is a non-profit organization created in 2014 by the city of Montreal to manage its bike-sharing system. The BIXI network has 6,200 bikes and 540 stations spread out across the areas of Montreal, Longueuil and Westmount. Much more than just a simple mode of transport, BIXI is now a great way to zip around the city whenever you want to go wherever your heart desires. BIXI Montréal, is an active mode of transportation contributing to the health and well-being of Montrealers. BIXI’s vision is to be recognized as a key player in Montreal’s public transit system.

## Analytics Requirements

BIXI ridership has been growing in recent years due to increased accessibility including the addition of 1,000 bikes, 80 stations and 2,214 docking points. In order to continue responding to growing demand more efficiently, and continue to offer high-quality customer service, BIXI needs improved revenue and cash flow forecasting to manage its capital expenditure program and to continue to invest in other initiatives that improve customer service and satisfaction.

Additionally, to ensure BIXI is serving the publics needs, it needs to monitor geographical ridership data to ensure that it is investing in bikes, stations, and docking points in those areas that are experiencing the highest growth in ridership across the Montreal area.

To achieve these goals, Group 10 is proposing the following analysis:

1. Weather – Ridership Correlation: Understanding the correlation between weather and the number of bike trips, or km travelled, will assist in forecasting revenues. As with many industries, we expect seasonality to affect ridership and, hence, revenue forecast to enable a more optimized capital expenditure management.
2. Station Ridership: Tracking ridership by station will help ensure that high demand areas have sufficient bikes and docking options or possibly relocating assets from low or underutilized areas to higher demand regions.

# Data Preparation & Analysis

In order for management to have confidence in the various forecasting tools proposed by Group 10, it is important to understand:

* The source and quality of the data used for our modeling
* How we procured this data
* The methods used to “clean” the data
* What tools or programming we conducted to analyse the data, along with associated challenges that management needs to be aware for future analytics needs
* The statistical methods used to analyse and interpret the dataset
* Conclusions derived from the analysis of this data.

# Data Source

Conducting the analysis requires three primary sources of data, specifically: BIXI ridership, BIXI station information, and Montreal weather data. BIXI ridership and station dataset can be readily downloaded from Kaggle which contains data from 2014 – 2017 Montreal weather data was downloaded from XXX.

The BIXI ridership dataset consists of over 1 million rows, for 2017, and 7 columns including start date, start station code, end date, end station code, duration of the ride (seconds), and membership status (i.e. member or non-member). The station dataset includes 545 stations and provides the station code, station name, latitude, and longitude. Weather data for temperature and humidity were obtained from XXX. Below is a top-level overview of the raw datasets.

Table 1: Ridership Raw Data

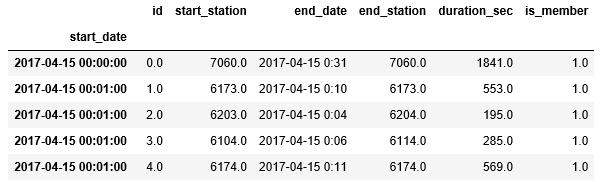
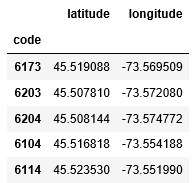


Table 2: Station Raw Date (Station Name Dropped)



# Data Quality & Cleaning Methods

## Quality

Ridership Data: The data quality for BIXI ridership was relatively good. The primary stages for cleaning the data include dropping three blank columns, converting date columns to a datetime data type, and setting the start date as the index for the dataset.

Station Data: Station data quality is also very good with no cleaning required. The only transformations performed on this dataset was to set the Station ID as the index to be used later to join with the ridership dataset.

Weather Data: Temperature and humidity data was also good but was provided in hourly increments whereas ridership data was provided by the minute. As such, the weather data was resampled to apply the temperature and humidity data for every minute within a given hour.

## Cleaning & Transformations

In reviewing the dataset, it was determined that a portion of rides have the same start and end stations. Since this initial analysis is focusing on the correlation between weather and distance travelled, this data was removed for the analysis and assigned to a dedicated dataframe.

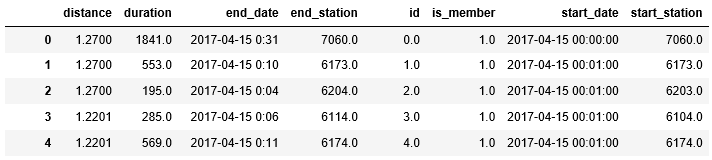


Table 3: Master Dataframe (Partial Sample)

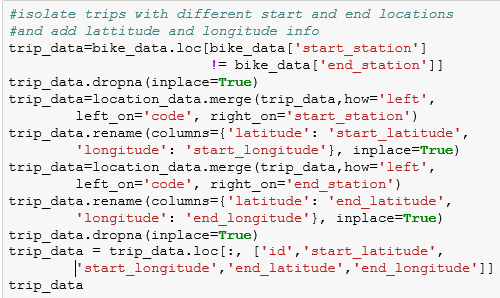


Figure 1: Code for removing trips with the same start and end stations

The rider and station datasets were joined using the Station Code (in the stations dataset) and Start Station (in the ridership dataset) columns. Once joined, each ride has start and end latitude and longitude coordinates assigned, allowing us to calculate the linear distance travelled, as detailed below. Additionally, weather data was joined the now master dataset so that ridership, distance, and weather data are all located in dataframe.

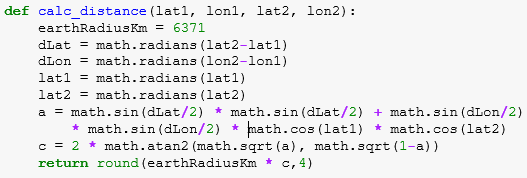
# Tools & Programs Used

Xxx

## Coding: Highlights

With the longitude and latitude data combined with the individual trips, it is possible to calculate the linear distance travelled using the formula shown below.

Figure 2: Method for Calculating Trip Distance



# Statistical Methods

Xxx

# Data Visualization

Xxx

# Analysis & Conclusions

Do not delete this Section Break, only delete this yellow highlight

1. TBD

xxx