



# Bike Sharing Program

07.09.2018

Rashi Nigam

Springboard Data Science Track

Fremont, CA 94539



## Overview

A **bicycle-sharing system**, is a service in which **bicycles** are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" and return it at another dock belong to the same system.

### Why the Need

- Increase personal mobility, providing people with better access to destinations throughout the City
- Integrate bike share as an extension of public transit network
- Develop an innovative transportation system that improves livability and economic competitiveness
- Reduce the environmental impact of transportation and help achieve goal of 'Go Green'
- Develop a system that serves users in minority and low-income communities and improves their access to key destinations, such as jobs and recreation
- safe mode of transportation that promotes active and healthy living
- Create a system that is financially sustainable, transparently operated, and accountable to the public.

History of Bike sharing systems go way back since year 1965 in Amsterdam however most of the major North American Systems started around 2010 and Capital Bikeshare - Washington, DC and Arlington, VA (1,600 bikes/191 stations) is among the most active systems. It is as well the client of this project. Capital Bikeshare has grown steadily, which has driven demand for more stations and bikes. Thus Client needed a research analysis to optimize their service & operations that

- Predict the Bike Rental volume/count
- Factors or features that influence Bike Rental Count



## Goals

Predict the Bike Rental volume from the dataset given by Capital Bike Sharing System.

Determine the factors or features that influence Bike Rental Count most.

The client is Capital Bikeshare System and this research to predict bike rental count will be useful to them in knowing:

- What features in the dataset influence the bike rental count
- When is the demand for bike share program maximum during the day, season, quarter or year.
- Does weather conditions like temperature, humidity, windspeed have any impact on the demand? If yes, then is it to advantage or adverse.
- Are bike users whether Registered or Casual drive Bike Rental Count? If yes, do they have similar influence on Bike Rental Count distribution?

## Data Wrangling

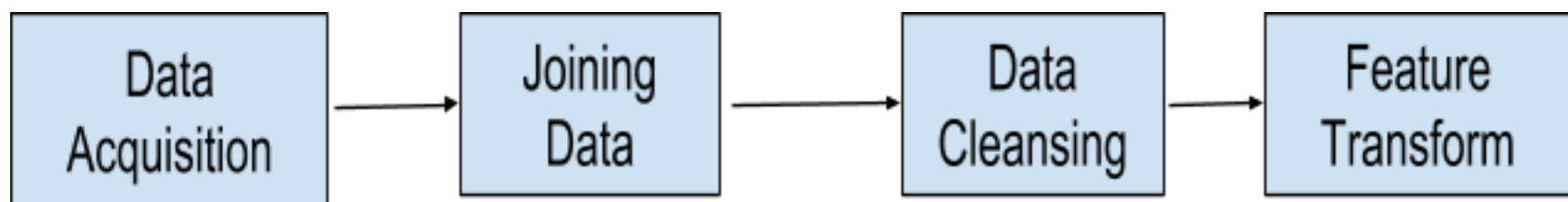
Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis.

Goals of Data Wrangling are:

- Reveal a “deeper intelligence” within your data, by gathering data from multiple sources
- Provide accurate, actionable data in the hands of business analysts in a timely matter
- Reduce the time spent collecting and organizing unruly data before it can be utilized
- Enable data scientists and analysts to focus on the analysis of data, rather than the wrangling
- Drive better decision-making skills by senior leaders in an organization

### Key steps in Data Wrangling





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.kaggle.com/c/bike-sharing-demand/data> and appear as below:

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

The Capital bikeshare datasets required data wrangling in terms of extracting dataset from Capital Bikeshare public shared repository followed by identifying meaningful dataset, renaming a few columns based on preference, feature transforming date timestamp to day, month, year; formatting the date and time columns to match with the weather data.

### 1. Renaming the columns as below:

'instant': 'rec\_id',

'dteday': 'datetime',

'holiday': 'is\_holiday',

'workingday': 'is\_workingday',

'weathersit': 'weather\_condition',

'hum': 'humidity',

'mnth': 'month',

'cnt': 'total\_count',







```
'hr': 'hour',
```

```
'yr': 'year'
```

**2. There were not any missing values to drop or replace. Type casting the attributes as 'datetime' or 'category' shown below**

```
stats['datetime'] = pd.to_datetime(stats.datetime)#dae time conversion
```

```
# categorical variables
```

```
stats['season'] = stats.season.astype('category')
```

```
stats['is_holiday'] = stats.is_holiday.astype('category')
```

```
stats['weekday'] = stats.weekday.astype('category')
```

```
stats['weather_condition'] = stats.weather_condition.astype('category')
```

```
stats['is_workingday'] = stats.is_workingday.astype('category')
```

```
stats['month'] = stats.month.astype('category')
```

```
stats['year'] = stats.year.astype('category')
```

```
stats['hour'] = stats.hour.astype('category')
```

























































































































