

# Bay Area Regional Transport Project





# Hello!

Zia Khan

[zia@thedevmasters.com](mailto:zia@thedevmasters.com)



# CRISP-DM

1

Business Objective

2

Data Understanding

3

Data Preparation

4

Modeling

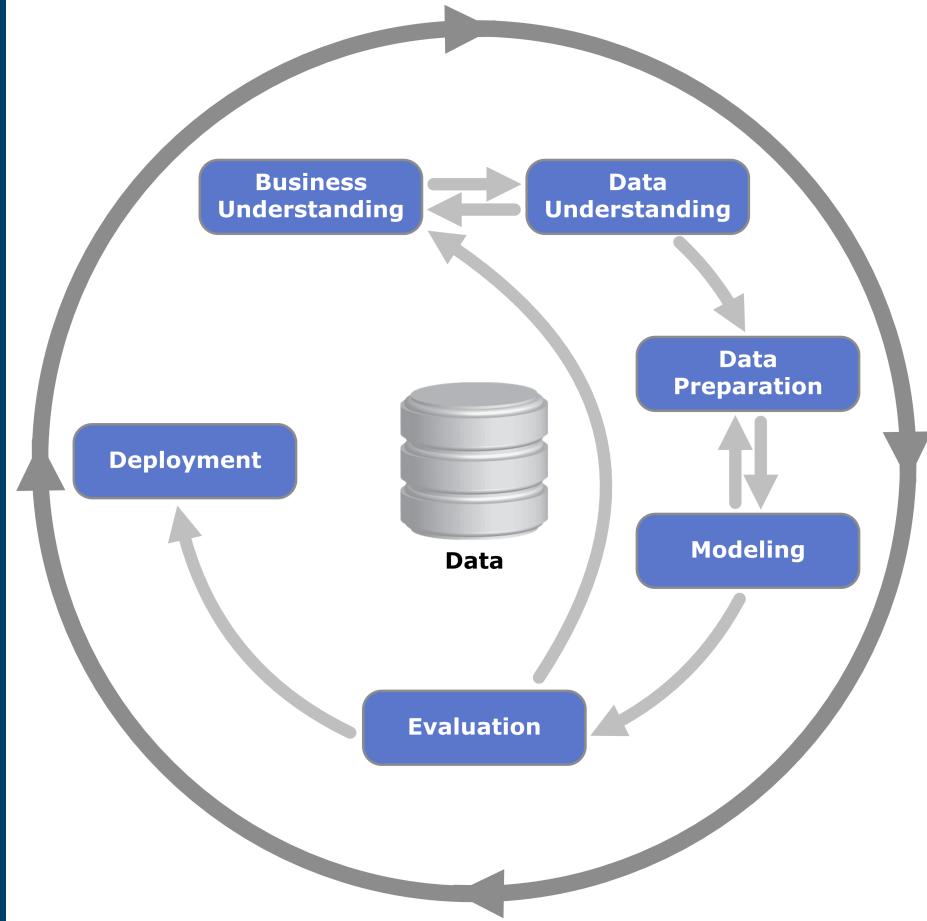
5



# CRISP-DM

## Overview

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment



**CRISP-DM**

**1**

**Business Objective**

**2**

**Data Understanding**

**3**

**Data Preparation**

**4**

**Modeling**

**5**



# 2. Objective

San Francisco BART is interested in developing a predictive analytics models to help predict ridership for staffing and operations



# Data Analytics Questions

- Which BART station is the busiest?
- What is the least popular BART route?
- When is the best time to go to SF from Berkeley if you want to find a seat?
- Which day of the week is the busiest?
- How many people take the BART late at night?



# Data Science Questions

- **Question A:** Compute the straight line distance between every station
- **Question B:** Build a model that can predict the number of people commuting to work by Bart between any 2 stations:
  - How would you approach that?
  - What additional data would you need
  - How would you want the city or Bart officials to use your data?



**CRISP-DM**

**1**

**Business Objective**

**2**

**Data Understanding**

**3**

**Data Preparation**

**4**

**Modeling**

**5**



# Data Understanding

	Origin	Destination	Throughput	DateTime
0	12TH	12TH	1	2016-01-01 00:00:00
1	12TH	16TH	1	2016-01-01 00:00:00
2	12TH	24TH	4	2016-01-01 00:00:00
3	12TH	ASHB	4	2016-01-01 00:00:00
4	12TH	BALB	2	2016-01-01 00:00:00

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9971582 entries, 0 to 9971581
Data columns (total 4 columns):
Origin          object
Destination     object
Throughput      int64
DateTime        object
dtypes: int64(1), object(3)
memory usage: 304.3+ MB
```

	Abbreviation	Description	Location	Name
0	12TH	1245 Broadway, Oakland CA 94612 12th St. ...	-122.271450,37.803768,0	12th St. Oakland City Center (12TH)
1	16TH	2000 Mission Street, San Francisco CA 94110<br...>	-122.419694,37.765062,0	16th St. Mission (16TH)
2	19TH	1900 Broadway, Oakland CA 94612 19th Stre...	-122.268602,37.808350,0	19th St. Oakland (19TH)
3	24TH	2800 Mission Street, San Francisco CA 94110<br...>	-122.418143,37.752470,0	24th St. Mission (24TH)



**CRISP-DM**

**1**

**Business Objective**

**2**

**Data Understanding**

**3**

**Data Preparation**

**4**

**Modeling**

**5**



# Data Preparation

- Append the two ridership files for 2016 and 2017
- Merge the station data file
- Answer data analytics questions



**CRISP-DM**

**1**

**Business Objective**

**2**

**Data Understanding**

**3**

**Data Preparation**

**4**

**Modeling**

**5**



# Modeling

- Build a model to predict ridership between two cities
- Answer the data science question





# Thanks!

Any questions?

[zia@thedevmasters.com](mailto:zia@thedevmasters.com)

