If you don't understand my GitHub Personal Project, this slide will describe the method and what I do in a nutshell.

My Personal Project with Online Data Set
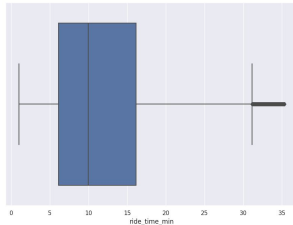
Jakkit Sida

Example Case 1: Process Data Anlyst "How Does a Bike-Share Navigate Speedy Success?"

1. Download File from Online Data Set and Import file to Google Colab (Python)

Divvy_Trips_2019_Q1
Divvy_Trips_2019_Q2
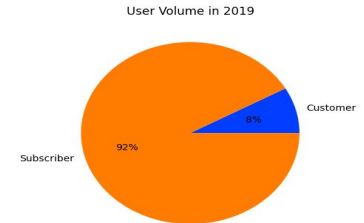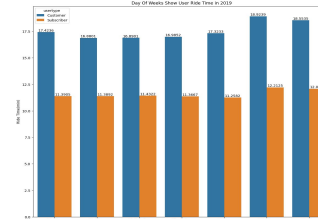Divvy_Trips_2019_Q3
Divvy_Trips_2019_Q4

2. Data Preprocessing (Cleaning, Tranformation)
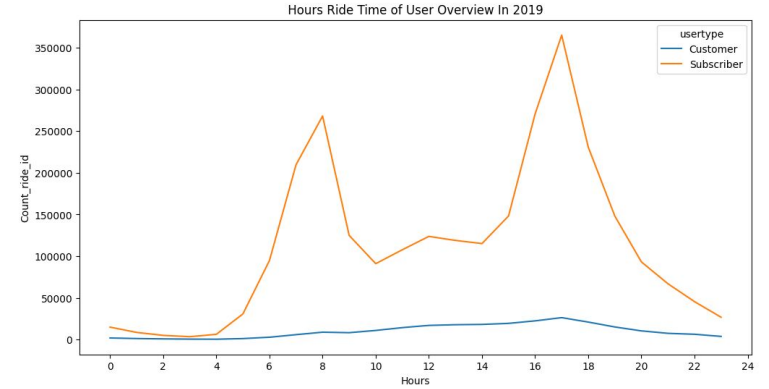
Combine_insight
merge_files
Remove_Outlier

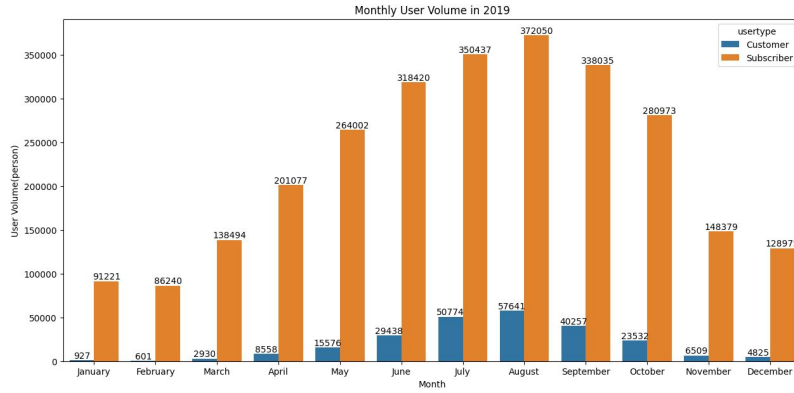4. Analyst & Conclusion and

Use extraction insight data with data drive discision methods

3. Create Visualization

Example Visualization & Insight

Monthly People Volume

Peak Time

Example Case 1: Process Data Science "Prediction Purchaing Behavior of Customer by ML"

1. Import Data and function library

2. Data Processing
(Clean, Transform) &
Analyst insight Data (Stastics, Graph)

5. Deployment Modeling
(Improve model, find suitable model)

3. Modeling (Sprit train/test data set,
Standard Suitable Model)

4. Evaluation Modeling
(Improve model, find suitable model)
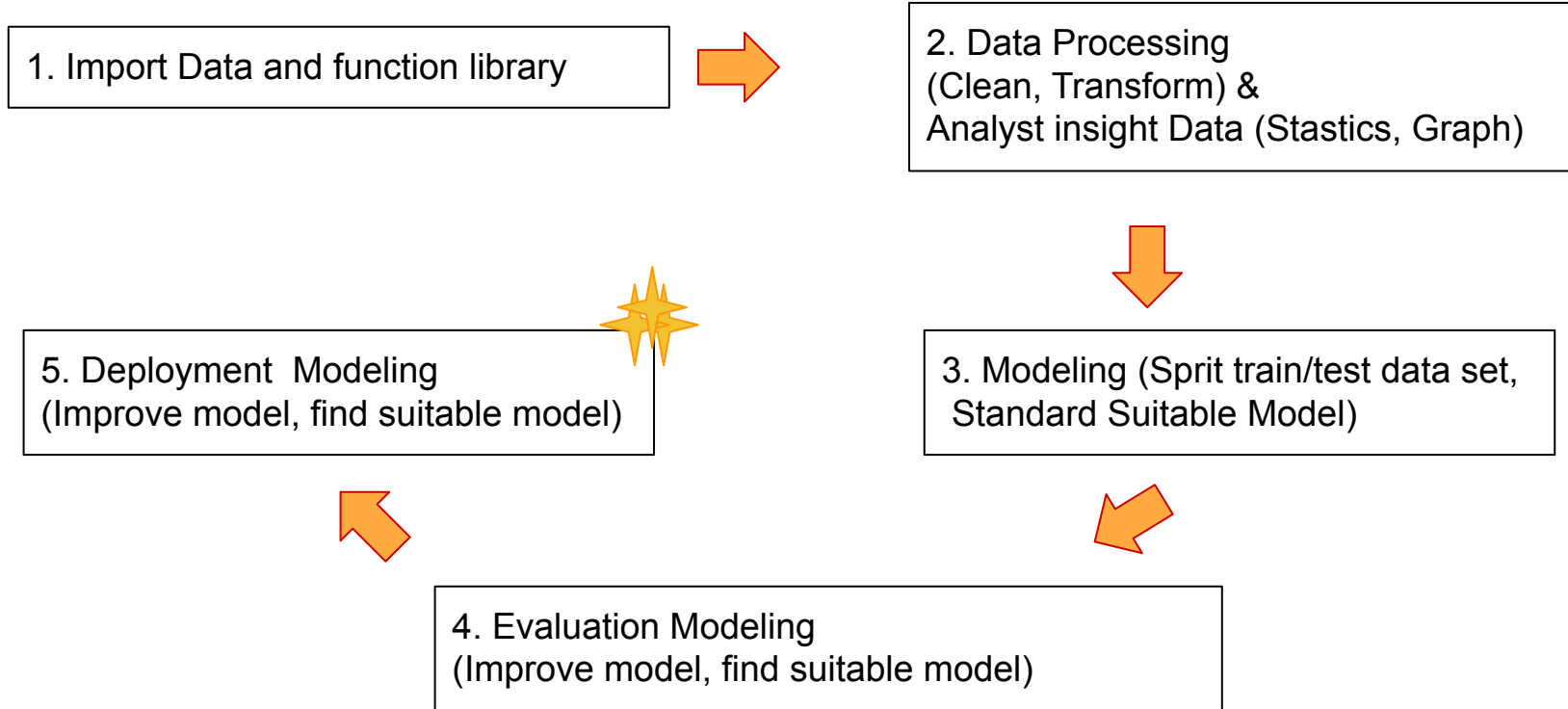
Example Model & Comparison

model	Score
KNeighborsClassifier	**0.8875**
RandomForestClassifier	0.8750
DecisionTreeClassifier	0.7875
LogisticRegression	0.5500

Trending Score

Then we get a prediction model that can predict whether a customer will buy or not with 87.5% accuracy.

It merely provides an overview; in practice, we must evaluate numerous factors and alter many times before we find an acceptable model.

```python
from sklearn.ensemble import RandomForestClassifier
modelRFC = RandomForestClassifier()
modelRFC.fit(X_train, y_train)
y_pred = modelRFC.predict(X_test)

evaluate_model_performance(y_test, y_pred)
```

```
0.875
[[37  7]
 [ 3 33]]
```

```python
acc_RFC = accuracy_score(y_test, y_pred)
print( 'Random Forest Classification: ', acc_RFC )
```