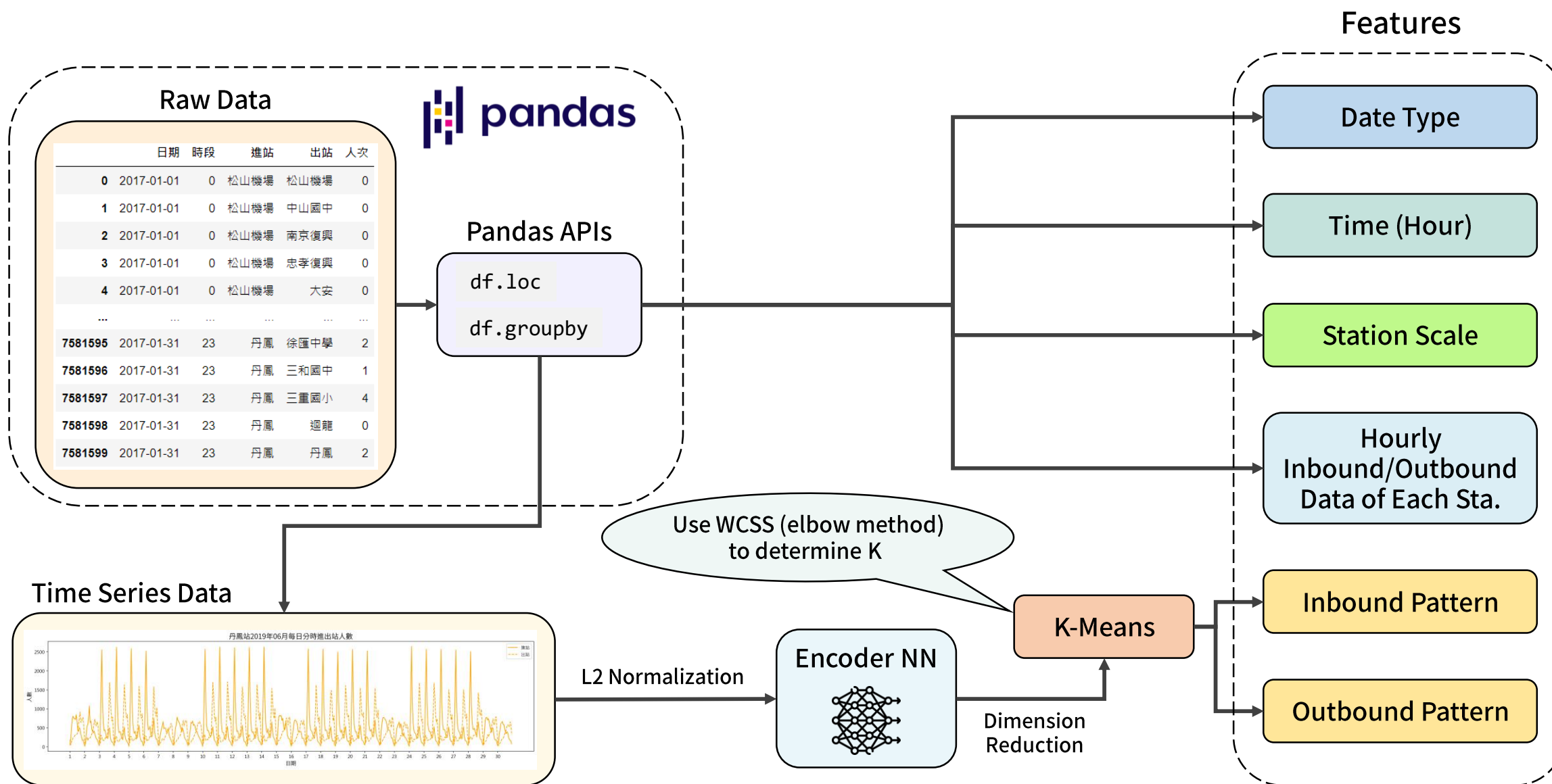# 🎯 Objective

Motivation

- To estimate the stream of people for each station in the next hour in real-time.
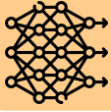
Scope

- Trip data for each station for 5 years (from 2017-01 to 2022-01).

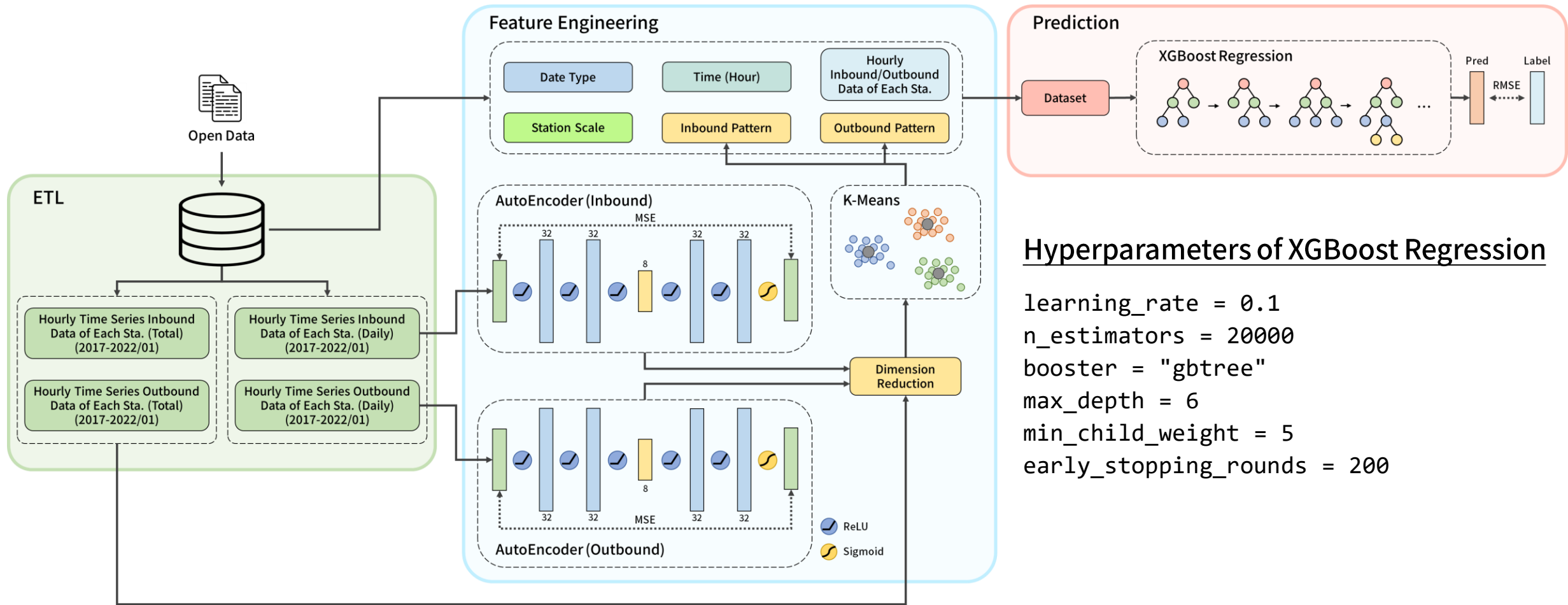- All the MRT stations are considered except the Circular Line.

Source: https://www.metro.taipei/cp.aspx?n=91974F2B13D997F1

Feature Engineering

Features

Raw Data

pandas

	日期	時段	進站	出站	人次
0	2017-01-01	0	松山機場	松山機場	0
1	2017-01-01	0	松山機場	中山國中	0
2	2017-01-01	0	松山機場	南京復興	0
3	2017-01-01	0	松山機場	忠孝復興	0
4	2017-01-01	0	松山機場	大安	0
...	...	...	...	...	...
7581595	2017-01-31	23	丹鳳	徐匯中學	2
7581596	2017-01-31	23	丹鳳	三和國中	1
7581597	2017-01-31	23	丹鳳	三重國小	4
7581598	2017-01-31	23	丹鳳	迴龍	0
7581599	2017-01-31	23	丹鳳	丹鳳	2

Pandas APIs

```
df.loc
df.groupby
```

Date Type

Time (Hour)

Station Scale

Hourly Inbound/Outbound Data of Each Sta.

Use WCSS (elbow method) to determine K

Time Series Data

丹鳳站2019年06月每日分時進出站人數

K-Means

Inbound Pattern

Outbound Pattern

L2 Normalization

Encoder NN

Dimension Reduction

Model Structure

ETL

Open Data

- Hourly Time Series Inbound Data of Each Sta. (Total) (2017-2022/01)
- Hourly Time Series Outbound Data of Each Sta. (Total) (2017-2022/01)
- Hourly Time Series Inbound Data of Each Sta. (Daily) (2017-2022/01)
- Hourly Time Series Outbound Data of Each Sta. (Daily) (2017-2022/01)

Feature Engineering

- Date Type
- Time (Hour)
- Hourly Inbound/Outbound Data of Each Sta.
- Station Scale
- Inbound Pattern
- Outbound Pattern

AutoEncoder (Inbound) — MSE — 32 32 8 32 32

AutoEncoder (Outbound) — 32 32 8 32 32 MSE

K-Means

Dimension Reduction

- ReLU
- Sigmoid

Prediction

XGBoost Regression

Dataset → ... → Pred — RMSE — Label

Hyperparameters of XGBoost Regression

```
learning_rate = 0.1
n_estimators = 20000
booster = "gbtree"
max_depth = 6
min_child_weight = 5
early_stopping_rounds = 200
```

Evaluation Strategy

Dataset
(3,810,348)

80% → Training
(3,048,278)

10% → Validation
(381,035)

10% → Testing
(381,035)

XGBoost Regression
(early stopping at 9656)

`model.score()`

MAPE

MAPE	Interpretation
< 10%	High accurate forecasting
10 – 20%	Good forecasting
20 – 50%	Reasonable forecasting
> 50%	Inaccurate forecasting

Results and its Possible Applications

Results
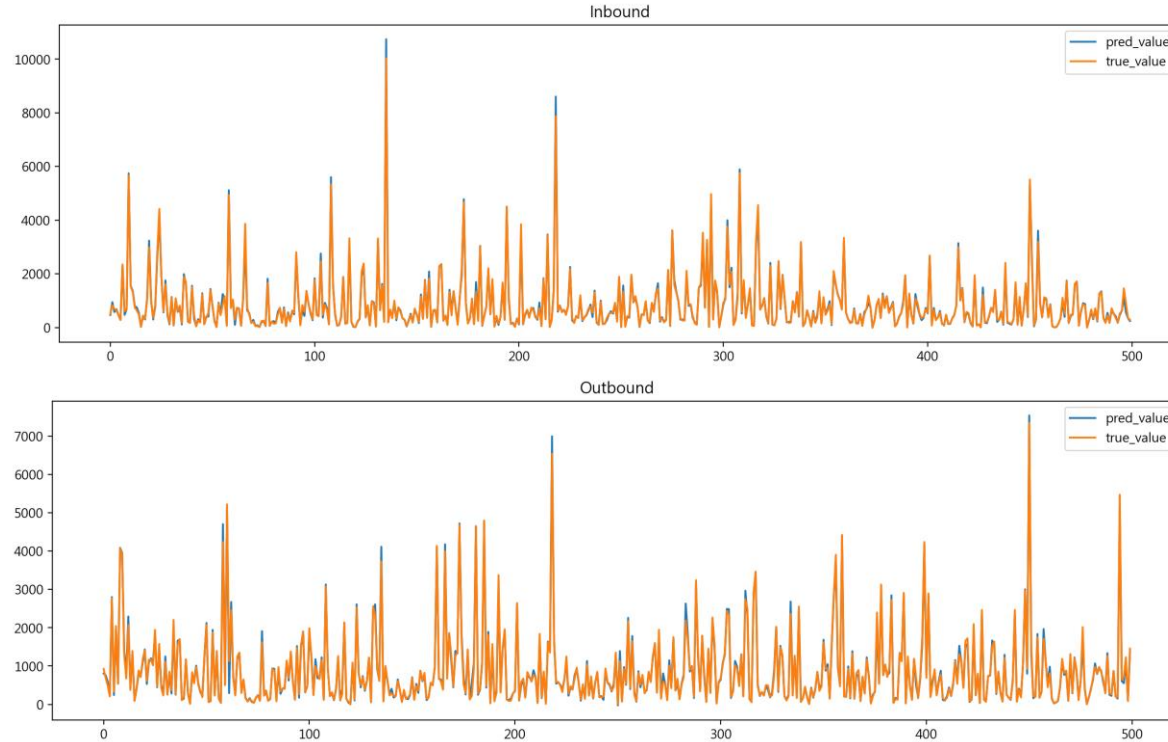
	Actual Inbound	Predict Inbound	MAPE
0	509	468.051697	8.04%
1	827	945.048767	14.27%
2	698	577.347656	17.29%
3	634	683.051147	7.74%
4	464	461.213928	0.60%
5	279	283.428833	1.59%
6	2354	2339.731689	0.61%
7	605	470.677399	22.20%
8	683	682.097473	0.13%
9	5677	5745.816406	1.21%
10	1520	1549.461426	1.94%
11	1323	1363.110107	3.03%
12	740	794.712341	7.39%
13	637	746.813354	17.24%
14	477	540.108643	13.23%
15	23	27.921343	21.40%
16	447	410.344696	8.20%
17	327	298.994598	8.56%
18	1008	1075.974243	6.74%
19	2997	3238.911865	8.07%
20	946	985.341187	4.16%

	Actual Outbound	Predict Outbound	MAPE
0	927	805.121033	13.15%
1	683	758.581116	11.07%
2	492	587.442261	19.40%
3	210	216.084381	2.90%
4	2767	2800.484619	1.21%
5	331	242.694992	26.68%
6	2045	1964.888062	3.92%
7	532	547.071533	2.83%
8	4070	4087.662598	0.43%
9	3900	3948.256836	1.24%
10	1289	1405.490723	9.04%
11	674	807.370911	19.79%
12	2072	2293.631104	10.70%
13	374	422.080200	12.86%
14	1394	1318.028809	5.45%
15	87	98.092339	12.75%
16	423	394.232300	6.80%
17	885	799.937012	9.61%
18	710	732.978027	3.24%
19	1178	1093.582642	7.17%
20	1409	1433.564087	1.74%

Application in the real world

- As an Early Warning System (EWS) to alert in real-time.
- As a Decision Support System (DSS) to assist operators in crowd management.