BIKE DEMAND PRESENTATION

Laïd ATTIA 24 November 2022

EXECUTIVE SUMMARY

- Bike demands are influenced by cities, available bicycles for rent, seasons, temperature, hour of the day and holidays
- Linear regression model is recommended to predict bike demand

Introduction

- Project is about how weather would affect bike-sharing demand in urban areas
- Data collection and sources
- Data exploration and analysis
- Data wrangling
- Data modelling
- Interactive Dashboard
- conclusion

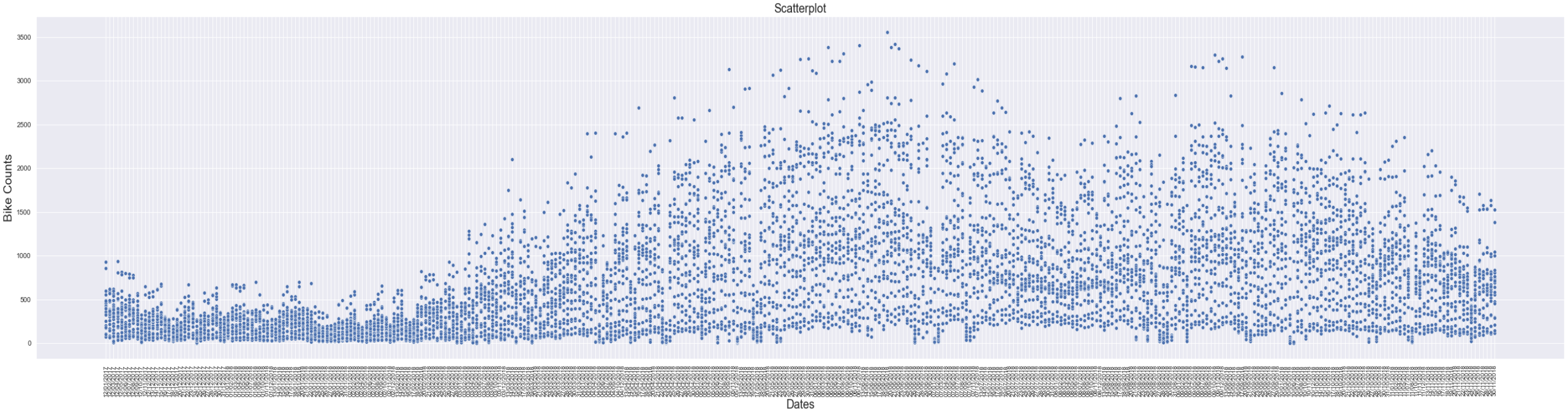
methodology

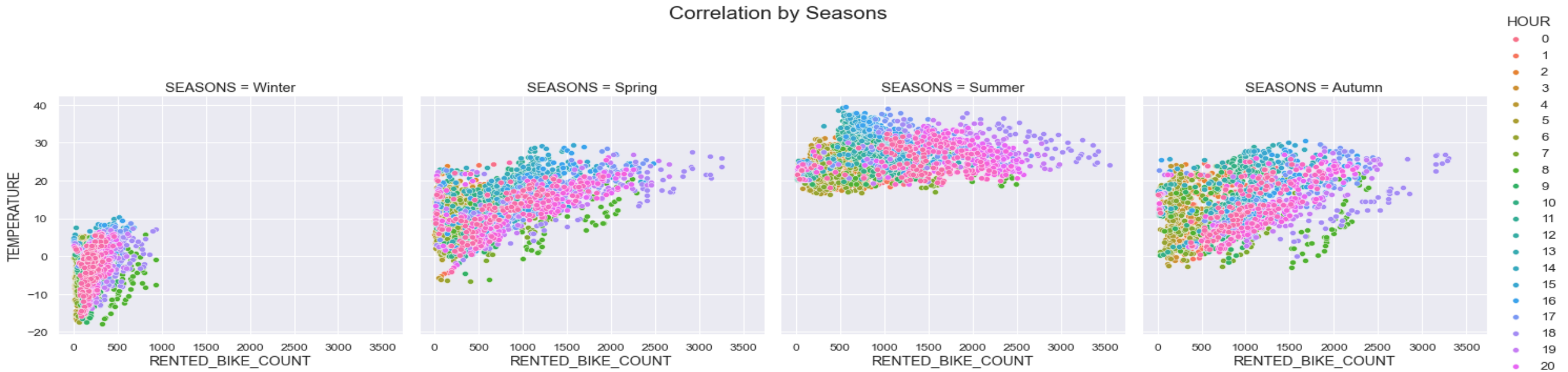
- Web scraping from Wikipedia (global bike sharing systems)
- 5 day weather forecasts for cities using openweather api
- Perform data wrangling on csv files
- Perform data exploration
- Perform data visualisation
- Predict Hourly Rented Bike Count using Basic Linear Regression Models
- Refine the Baseline Regression Models

Exploratory data analysis

- Dataset has 8465 observations for seoul
- Date range is from 1 dec 2017 to 30 Nov 2018 (one year)
- Highest bike count is 3556 on 19 June 2018
- Summer has most bike rentals, winter the least
- Demand varies from city to city

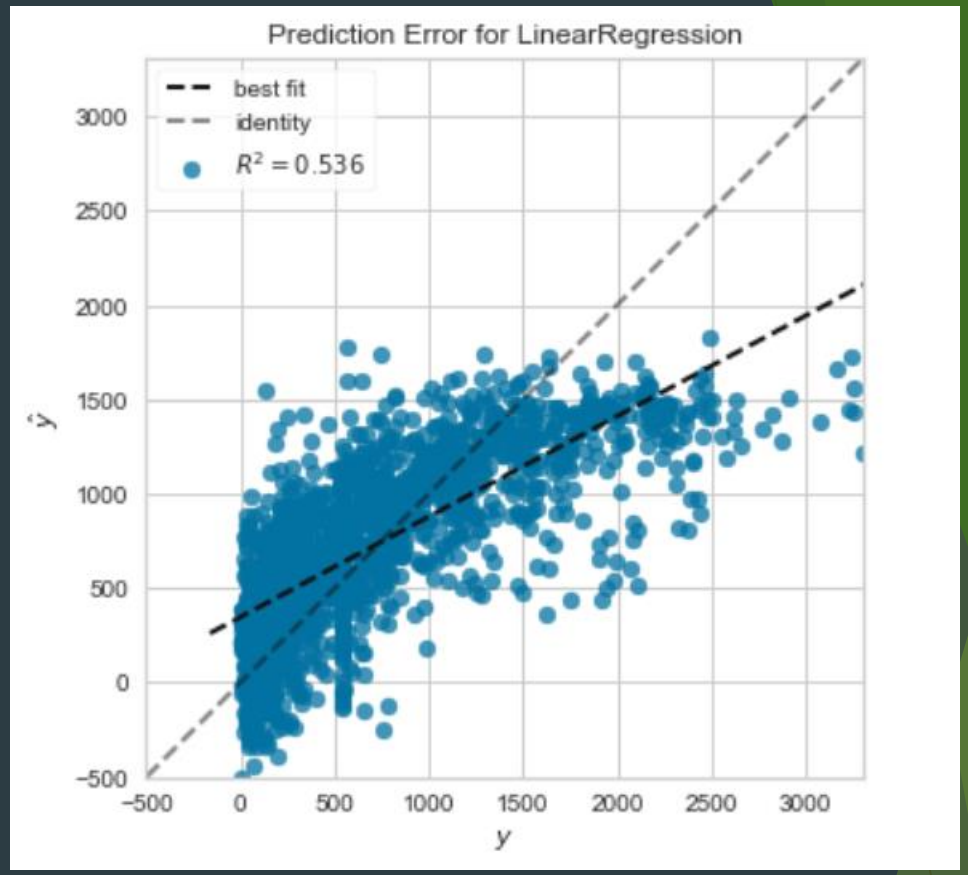
data visualization





Predictive analysis

- baseline regression model
- improving the model



	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lr	Linear Regression	283.0143	142660.9000	377.6294	0.6422	0.8924	1.5912	1.1580
ridge	Ridge Regression	282.9418	142645.2344	377.6091	0.6422	0.8941	1.5886	0.0140
lar	Least Angle Regression	283.0429	142767.5492	377.7662	0.6420	0.8922	1.5900	0.0160
lasso	Lasso Regression	283.3254	143644.9844	378.9253	0.6398	0.8928	1.5563	0.0400
llar	Lasso Least Angle Regression	354.6801	227398.6736	476.6986	0.4301	0.9387	1.9962	0.0160
en	Elastic Net	359.0337	235343.4406	484.8899	0.4107	0.9293	1.7635	0.0160

Interactive Dashboard

Bike-Sharing Demand Prediction App

Bike Sharing data

	DATE	RENTED_BIKE_COUNT	HOUR	TEMPERATURE	HUMIDITY	WI
0	2017-01-12T00:00:00+08…	254	0	-5.2000	37	
1	2017-01-12T00:00:00+08…	204	1	-5.5000	38	
2	2017-01-12T00:00:00+08…	173	2	-6	39	
3	2017-01-12T00:00:00+08…	107	3	-6.2000	40	
4	2017-01-12T00:00:00+08…	78	4	-6	36	

Selected Cities Data

	CITY	CITY_ASCII	lat	lon	COUNTRY	ISO2	ISO3	
0	Seoul	Seoul	37.5833	127	Korea, South	KR	KOR	
1	New York	New York	40.6943	-73.9249	United States	US	USA	
2	Paris	Paris	48.8566	2.3522	France	FR	FRA	Î
3	London	London	51.5072	-0.1275	United Kingdom	GB	GBR	Lon
4	Suzhou	Suzhou	31.3040	120.6164	China	CN	CHN	

City:

All

Select a button

Selections

- (●) Basic max bike prediction
- ( ) A static temperature trend line
- ( ) An interactive bike-sharing demand prediction trend line
- ( ) A static humidity and bike-sharing demand prediction correlation plot

conclusion

- Bike demands are influenced by cities, available bicycles for rent, seasons, temperature, hour of the day and holidays
- Linear regression model is recommended to predict bike demand
- Explore more sources on bikes demand online for analysis
- Use machine learning algorithms (decision tree, random forest, xgboost, gradient boost)
- Potential use for marketing and business expansions

appendix

Python code sample

```
fig = plt.figure(figsize=(50,10))
sns.scatterplot(x=df.DATE,y=df.RENTED_BIKE_COUNT,data=df, estimator=None)
plt.title("Scatterplot", fontsize=20)
plt.xticks(rotation=90, fontsize=10)
plt.xlabel("Dates", fontsize=20)
plt.ylabel("Bike Counts", fontsize=20)
plt.show()
```

Sql code sample

```
db = sqlite3.connect("bike.db") #Create connection
```

```
db
```

```
<sqlite3.Connection at 0x21b9cc21030>
```

```
cursor = db.cursor()
```

```
pd.read_sql_query("Select * FROM sqlite_master", db)
```

	type	name	tbl_name	rootpage	sql
0	table	SEOUL_BIKE_SHARING	SEOUL_BIKE_SHARING	2	CREATE TABLE "SEOUL_BIKE_SHARING" (\n"DATE" TE...
1	table	CITIES_WEATHER_FORECAST	CITIES_WEATHER_FORECAST	173	CREATE TABLE "CITIES_WEATHER_FORECAST" (\n"CIT...
2	table	BIKE_SHARING_SYSTEMS	BIKE_SHARING_SYSTEMS	178	CREATE TABLE "BIKE_SHARING_SYSTEMS" (\n"COUNTR...
3	table	WORLD_CITIES	WORLD_CITIES	184	CREATE TABLE "WORLD_CITIES" (\n"CITY" TEXT,\n ...

```
db.close()
```