BIKE DEMAND PRESENTATION:

IBM DATA ANALYTICS WITH EXCEL & R PROFESSIONAL CERTIFICATE FINAL ASSIGNMENT

AUTHOR: JOHN SMITH

LAST UPDATED: 2/20/2023

EXECUTIVE SUMMARY

- Bike Demand Influenced by Number of Variables
  - Cities, Available Bicycles for Rent, Seasons, Temperature, Hour of the Day, Holidays
- Exploratory Data Analysis Suggests Linear Regression Model
  - Given the Characteristics & Distribution of the Dataset, LRM is Most Appropriate Statistical Model
    - Utilized to Predict Bike Demand

INTRODUCTION – DEFINING BIKE DEMAND PROBLEM & GOALS OF PRESENTATION

- **Project Purpose:** Explore How Weather & Other Factors Affect Bike-Sharing Demand in Urban Areas
- **Presentation Organization:**
  - Data Collection & Sources
  - Data Exploration & Analysis
  - Data Wrangling Methods
  - Data Modeling Methods
  - Design & Implementation of Interactive Dashboard
  - Concluding Remarks

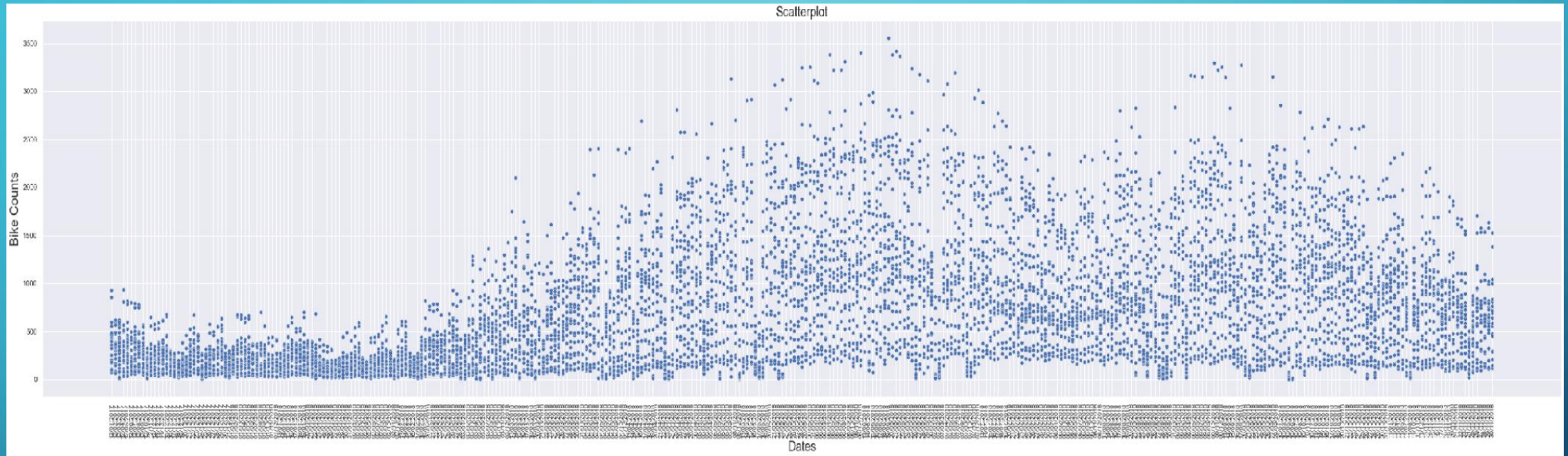
METHODOLOGY – REFINING THE BIKE DEMAND PROBLEM

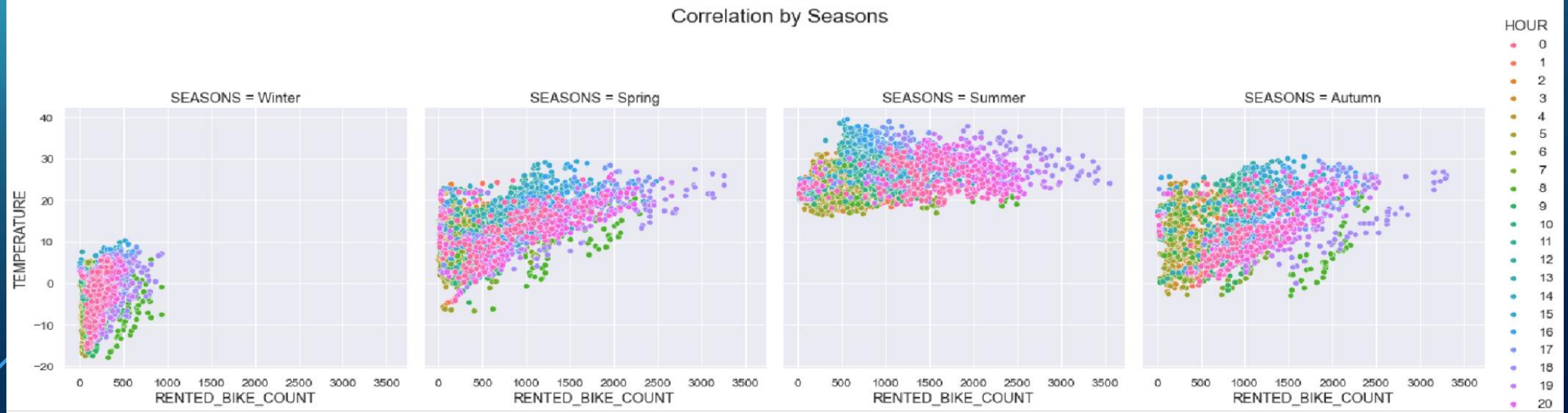
- **Data Sources:** Web Scraping via Wikipedia Entry & OpenWeather API
  - Global Bike Sharing Systems Dataset + 5 Day Weather Forecasts for Relevant Cities
- Perform Data Wrangling on Relevant CSV Files Appropriate for Analysis
- Perform Data Exploration Techniques to Gain Fundamental Understanding on Nature of Dataset
- Perform Data Visualization Techniques to Gain Insights from Dataset
- Build Model to Predict Hourly Rented Bike Count
  - Linear Regression Model
- Refine the Regression Model Employed
- Design Interactive Dashboard

EXPLORATORY DATA ANALYSIS – UNDERSTANDING THE DATASET

- Dataset has 8,465 Observations for Seoul, South Korea
- Dataset Range is from December 1, 2017 to November 30, 2018
  - ~1 Year of Latest Time-Series Data Available
- The Highest Bike Count is 3,556
  - Event Occurred on June 19. 2018
- There is a Seasonality Factor to Bike Demand
  - Bike Rental Demand is Highest in Summer & Lowest in Winter
- Bike Demand Varies Wildly from City to City Across the Globe
  - Cultural Factors & Quality of Infrastructure Likely Contribute to Variation

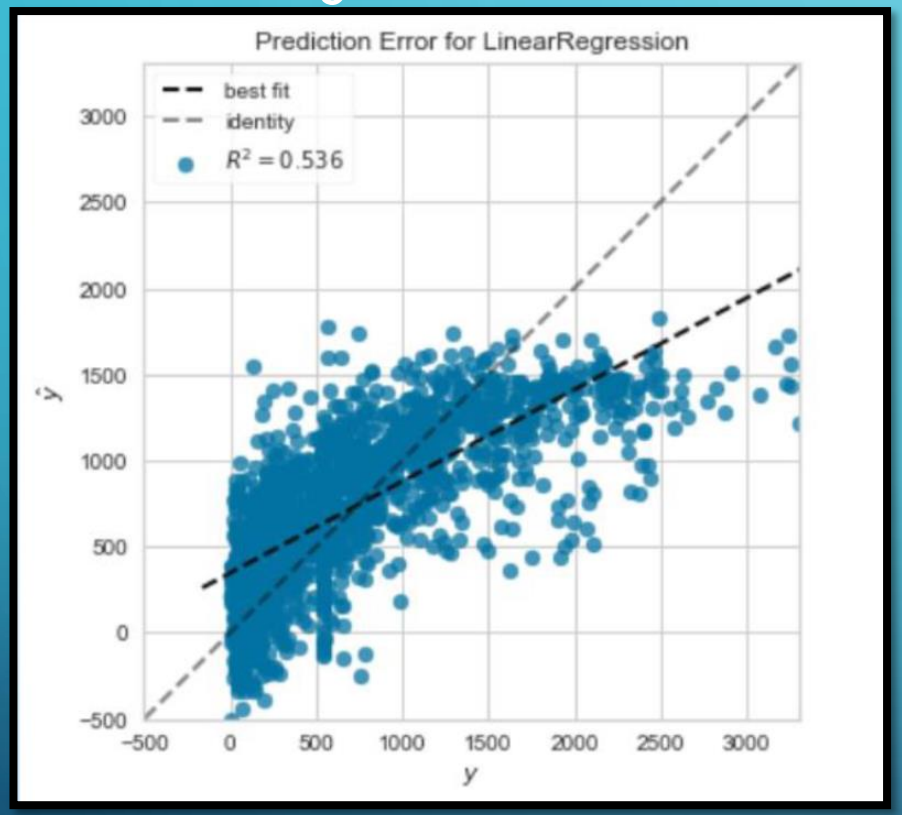
DATA VISUALIZATION – GAINING INSIGHTS FROM THE DATASET





DATA MODELING – EVALUATION OF MODELS TO PREDICT HOURLY RENTED BIKE COUNTS

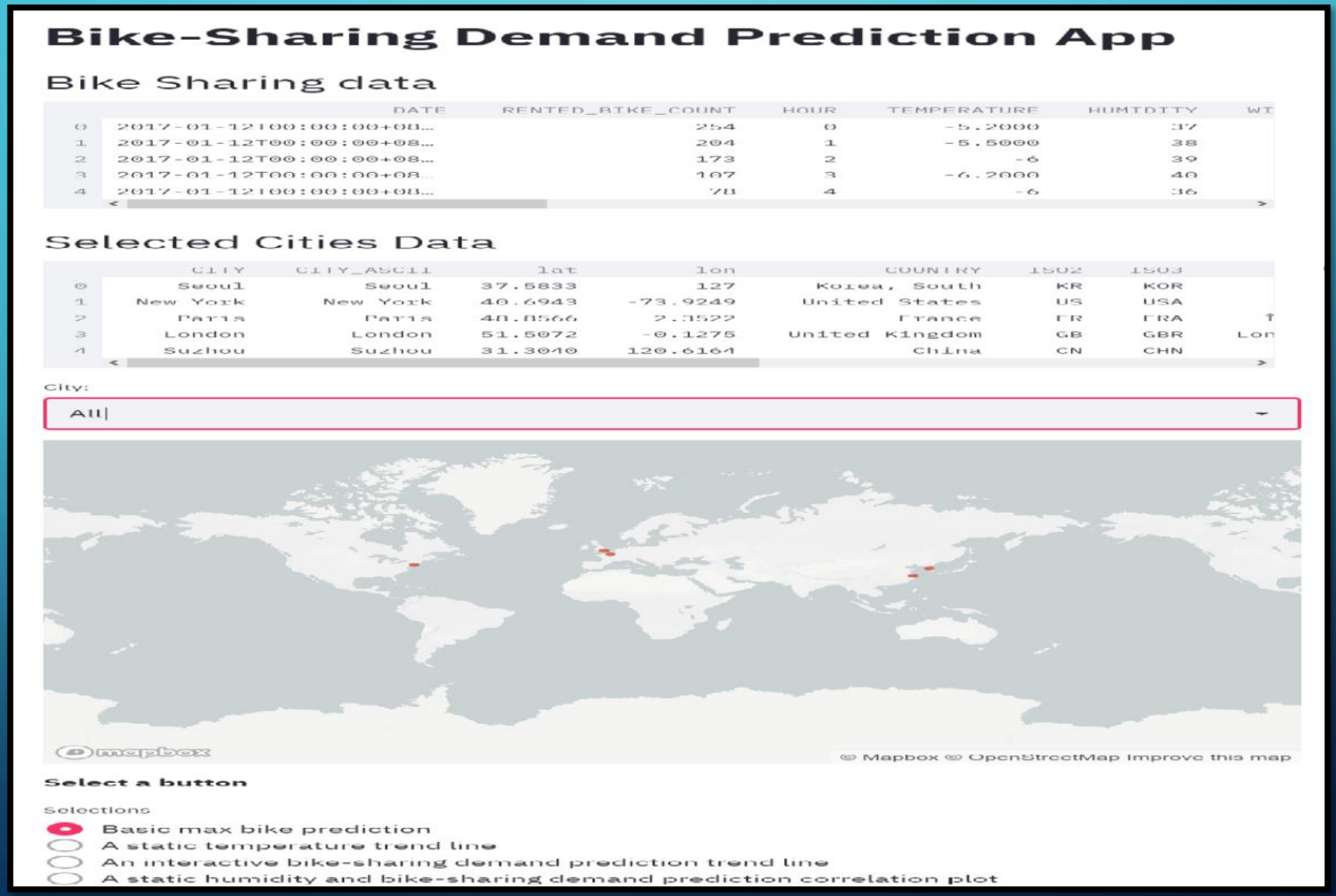
Baseline Regression Model



Improving Regression 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

BUILDING INTERACTIVE DASHBOARD – BIKE SHARING DEMAND PREDICTION APPLICATION



CONCLUSIVE REMARKS – TELLING THE STORY OF BIKE DEMAND & RELATED DATA

- **Bike Demands:**
  - Influenced by Cities, Available Number of Bicycles for Rent, Seasons, Temperature, Hour of the Day, and Holidays
- Linear Regression Most Appropriate Statistical Model to Predict Bike Demand
- Insights from Analysis Provide Potential Avenue for Marketing Programs & Business Expansionary Programs into Different Territories
- **Further Research:**
  - More Sources on Bikes Demand Available Online for Analysis
  - Implementation of Machine Learning Algorithms to Refine Statistical Models
    - Decision Trees, Random Forest, XGBOOST, and Gradient Boost

APPENDIX – PROVIDING SAMPLES OF CODE FROM ANALYSIS PROCEDURES

- 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()
```