CBD 2214: ASSIGNMENT 1

TOPIC: Power consumption of Tetouan city Data Set

Group 6

Dataset Description:

This dataset is related to power consumption of three different distribution networks of Tetouan city which is located in north Morocco. The data is being recorded every 10 minutes interval between January 1, 2017 to December 31, 2017.

Attribute Information:

Data was collected from January 1, 2017 to December 21, 2017 in 3 zones of Tetouan City simultaneously within a 10 minutes interval.

Dependent:

1. Date Time: Each ten minutes.

2. Temperature: Weather Temperature of Tetouan city.

3. Humidity: Weather Humidity of Tetouan city.

4. Wind Speed of Tetouan city.

5. general diffuse flows

6. diffuse flows

Independent:

7. power consumption of zone 1 of Tetouan city.

8. power consumption of zone 2 of Tetouan city.

9. power consumption of zone 3 of Tetouan city.

Analysis:

The Data set is clean and does not have any empty values which is a good thing, but by breaking the DateTime column into three columns (Hours, Weeks, Months) the results were showing a better accuracy and it was more readable.

	DateTime	Temperature	Humidity	Wind Speed	general diffuse flows	diffuse flows	Zone 1 Power Consumption	Zone 2 Power Consumption	Zone 3 Power Consumption
52411	12/30/2017 23:10	7.010	72.4	0.080	0.040	0.096	31160.45627	26857.31820	14780.31212
52412	12/30/2017 23:20	6.947	72.6	0.082	0.051	0.093	30430.41825	26124.57809	14428.81152
52413	12/30/2017 23:30	6.900	72.8	0.086	0.084	0.074	29590.87452	25277.69254	13806.48259
52414	12/30/2017 23:40	6.758	73.0	0.080	0.066	0.089	28958.17490	24692.23688	13512.60504
52415	12/30/2017 23:50	6.580	74.1	0.081	0.062	0.111	28349.80989	24055.23167	13345.49820

```
df["DateTime"]=pd.to_datetime(df["DateTime"])
```

```
df["Month"]=df["DateTime"].dt.month
df["Week"]=df["DateTime"].dt.weekday
df["hour"]=df["DateTime"].dt.hour
```

```
df.head(200)
```

	DateTime	Temperature	Humidity	Wind Speed	general diffuse flows	diffuse flows	Zone 1 Power Consumption	Zone 2 Power Consumption	Zone 3 Power Consumption	Month	Week	hour
0	2017-01-01 00:00:00	6.559	73.8	0.083	0.051	0.119	34055.69620	16128.87538	20240.96386	1	6	0
1	2017-01-01 00:10:00	6.414	74.5	0.083	0.070	0.085	29814.68354	19375.07599	20131.08434	1	6	0
2	2017-01-01 00:20:00	6.313	74.5	0.080	0.062	0.100	29128.10127	19006.68693	19668.43373	1	6	0
3	2017-01-01 00:30:00	6.121	75.0	0.083	0.091	0.096	28228.86076	18361.09422	18899.27711	1	6	0
4	2017-01-01 00:40:00	5.921	75.7	0.081	0.048	0.085	27335.69620	17872.34043	18442.40964	1	6	0
...	...	...	...	...	...	...	...	...	...	...	...	...
195	2017-01-02 08:30:00	8.710	83.9	0.081	6.657	6.472	24881.01266	16737.99392	12248.67470	1	0	8
196	2017-01-02 08:40:00	8.520	85.1	0.081	13.020	11.560	25190.88608	16891.18541	12404.81928	1	0	8
197	2017-01-02 08:50:00	8.430	85.4	0.079	39.100	19.490	25531.13924	17135.56231	12358.55422	1	0	8
198	2017-01-02 09:00:00	8.440	85.9	0.080	64.110	24.490	25980.75949	17248.63222	12636.14458	1	0	9
199	2017-01-02 09:10:00	8.690	86.7	0.079	92.300	27.880	23495.69620	17668.08511	13110.36145	1	0	9
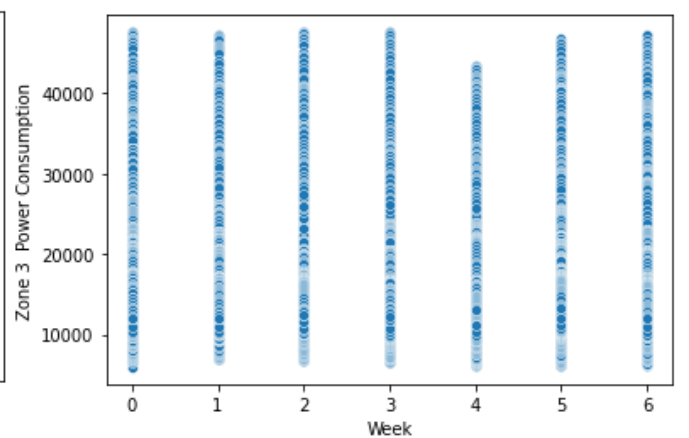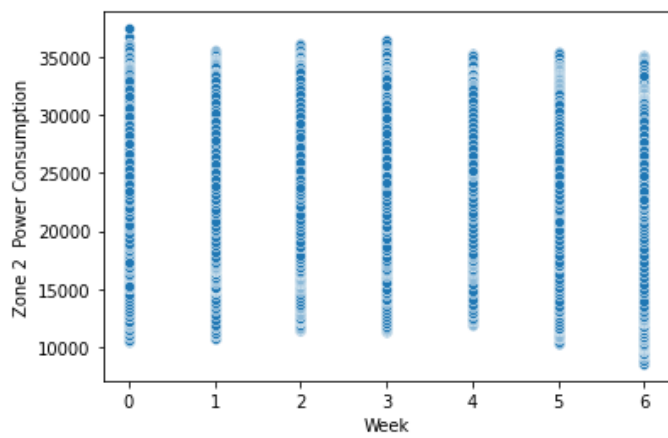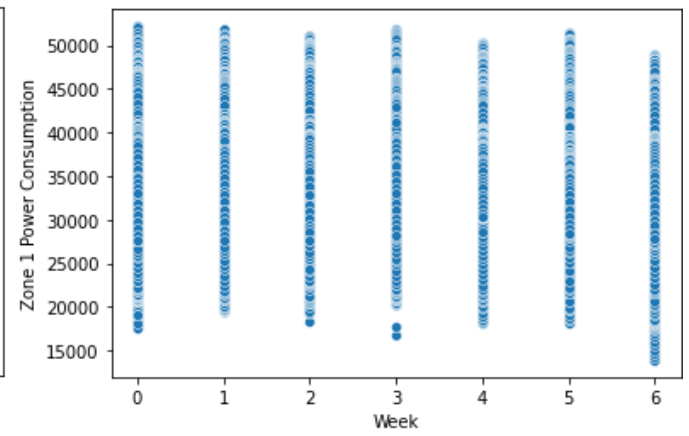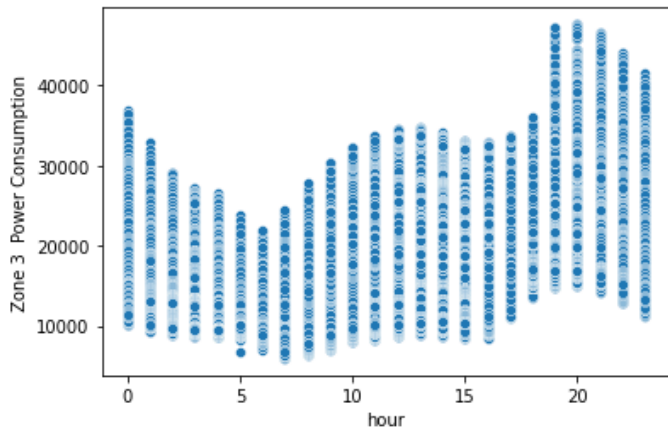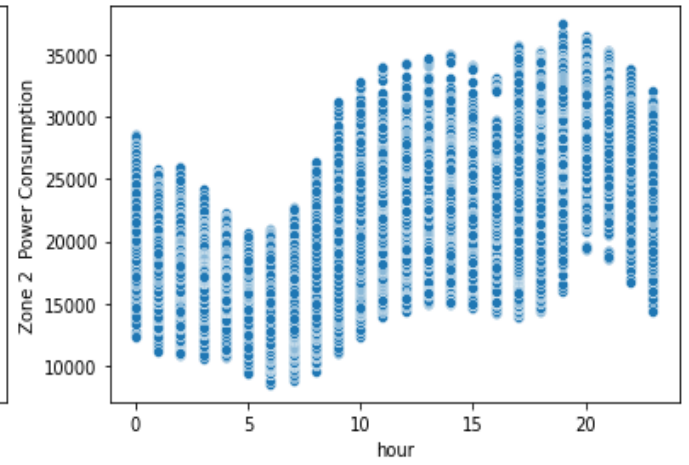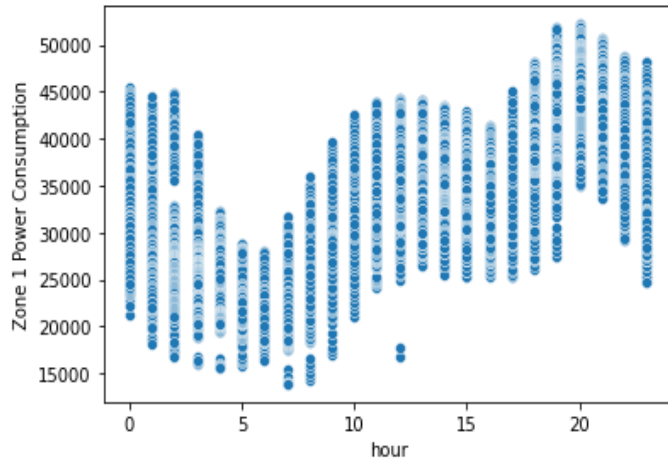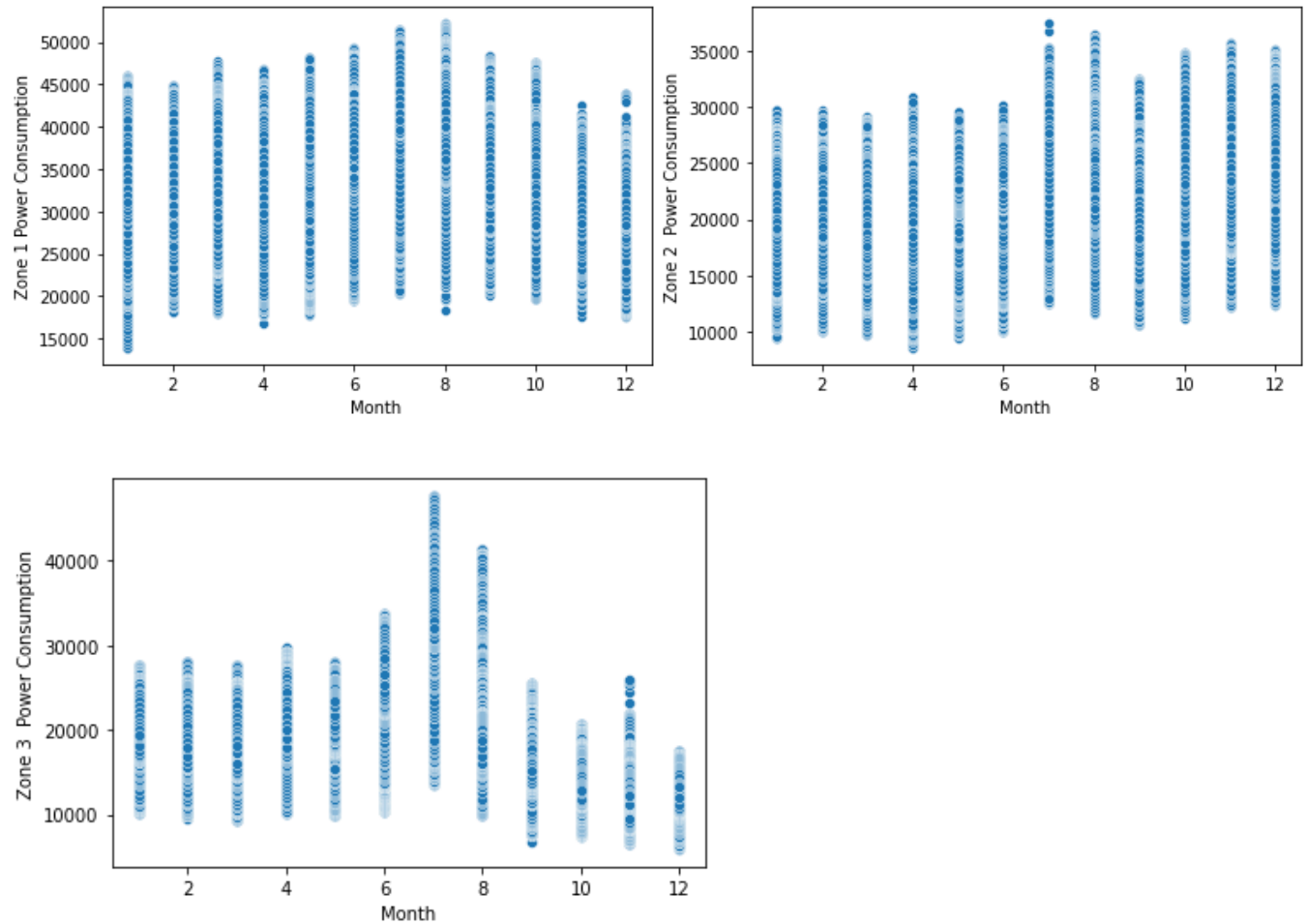
Based on what we can see in the charts we can see that between 20-24 hours is the most

consumption in all zones and in July is the most used power consumption as well in all zones.

Using multiple regression's algorithms we tried to predict the power consumption of the three

zones using all columns except the DateTime because it has been broken into 3 columns.

Algorithm	Split	Score	RMSE
Linear Regression	75 % was used for training and the rest 25 % was kept for testing the model.	0.6149510434725897	3948.051
SVR	75 % was used for training and the rest 25 % was kept for testing the model.	0.448033584003656	0.371
Random Forest Regression	80 % was used for training and the rest 20 % was kept for testing the model.	0.9808149990761602	873.575

Conclusion:

Random forest regression gave the best results with 98% accuracy where the highest accuracy

between the other algorithms was 61%.

Reference:

Scikit-learn.org., sklearn.ensemble.RandomForestRegressor

 https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Salam, A., & El Hibaoui, A. (2018, December). Comparison of Machine Learning Algorithms

for the Power Consumption Prediction:-Case Study of Tetouan cityâ€". In 2018 6th International

Renewable and Sustainable Energy Conference (IRSEC) (pp. 1-5). IEEE.â€