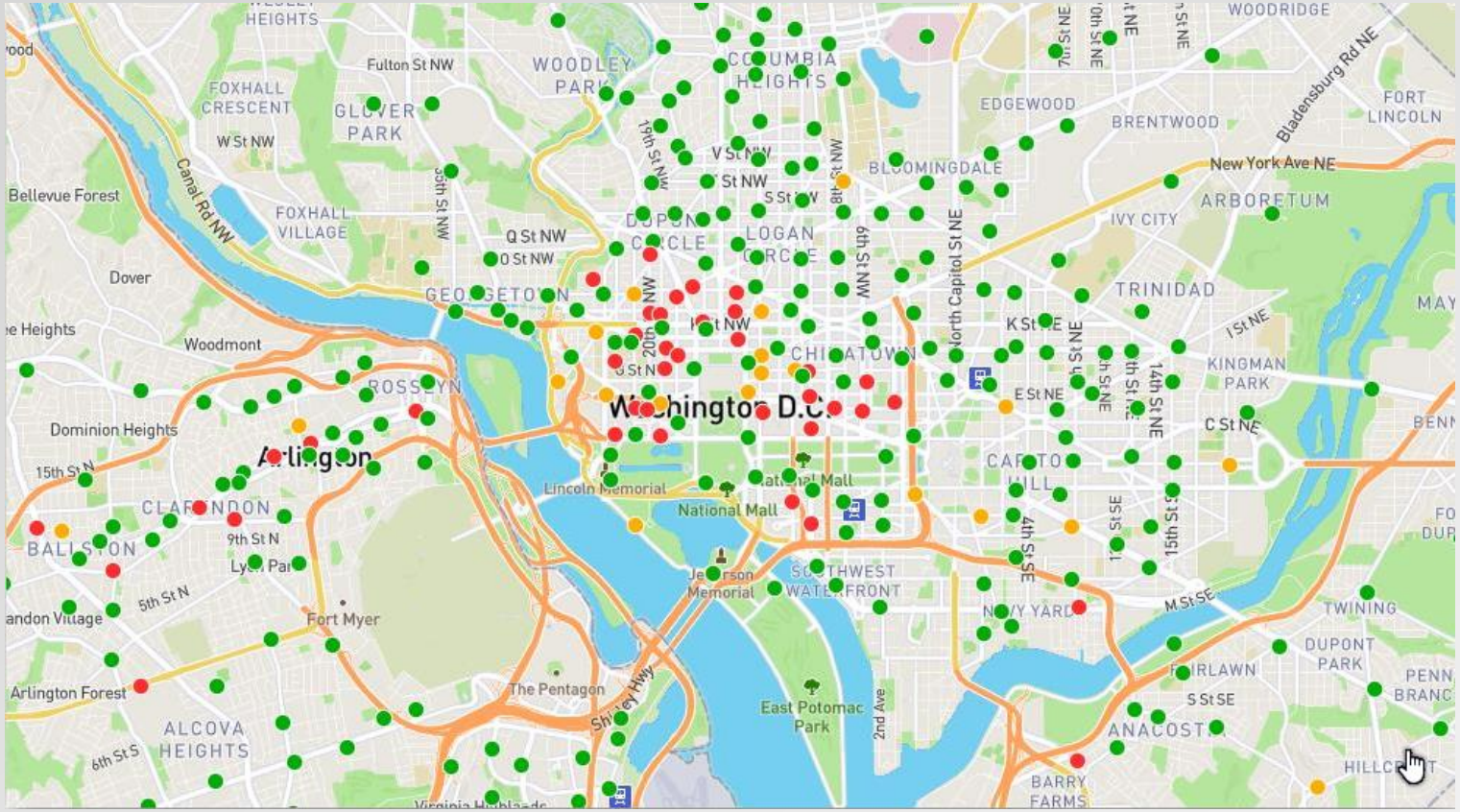
BIKE SHARING DEMAND

Forecast use of a city bikeshare system in Washington D.C.

Project Goal:

- **Prediction of the total number of bikes in Washington D.C. that will be rented in each hour.**

Hypothesis: The bike sharing are highly related with the time of the day, season and weather conditions.



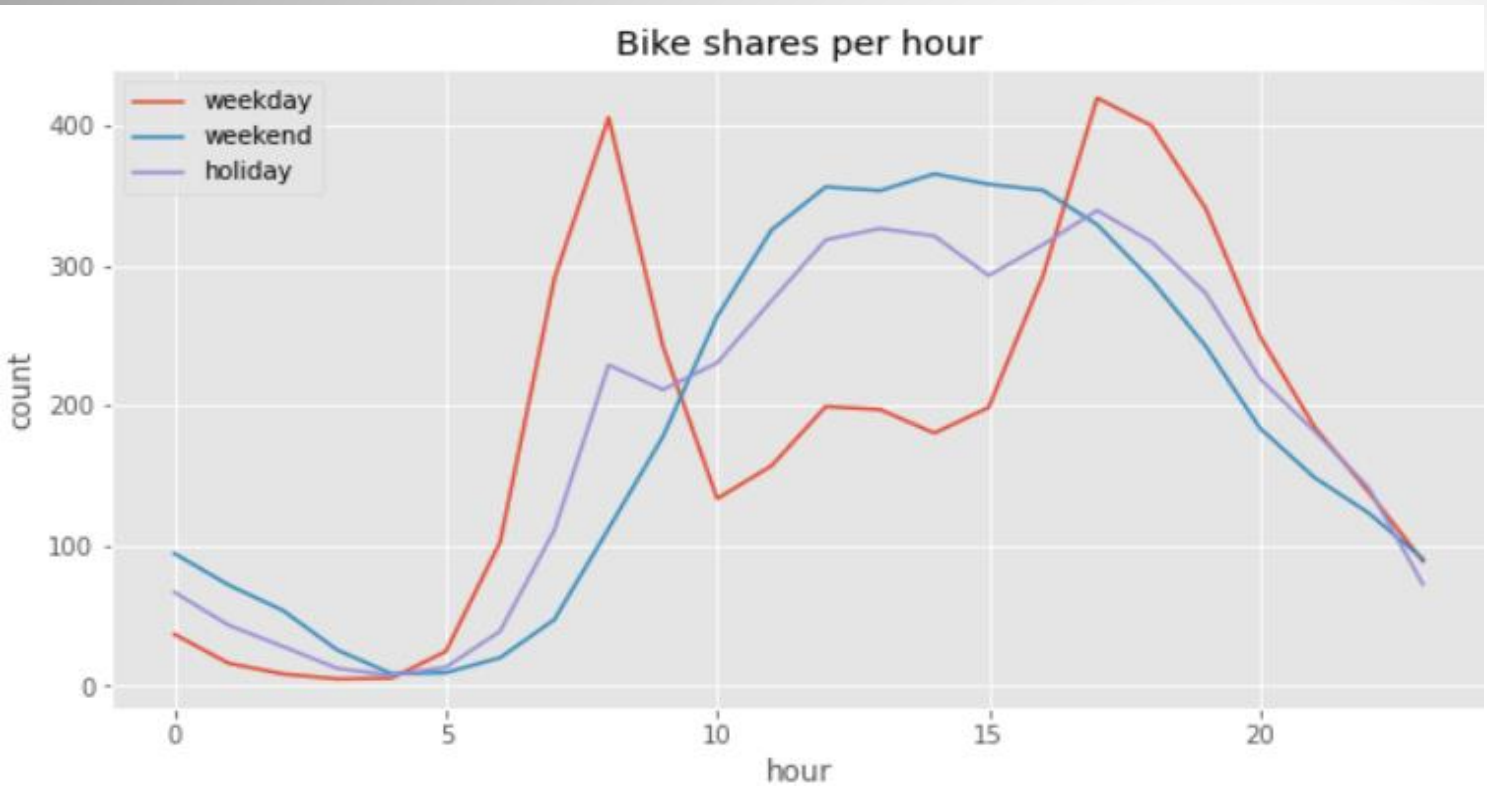
https://bikeshare.cc/guide/list-of-capital-bikeshare-station-maps/

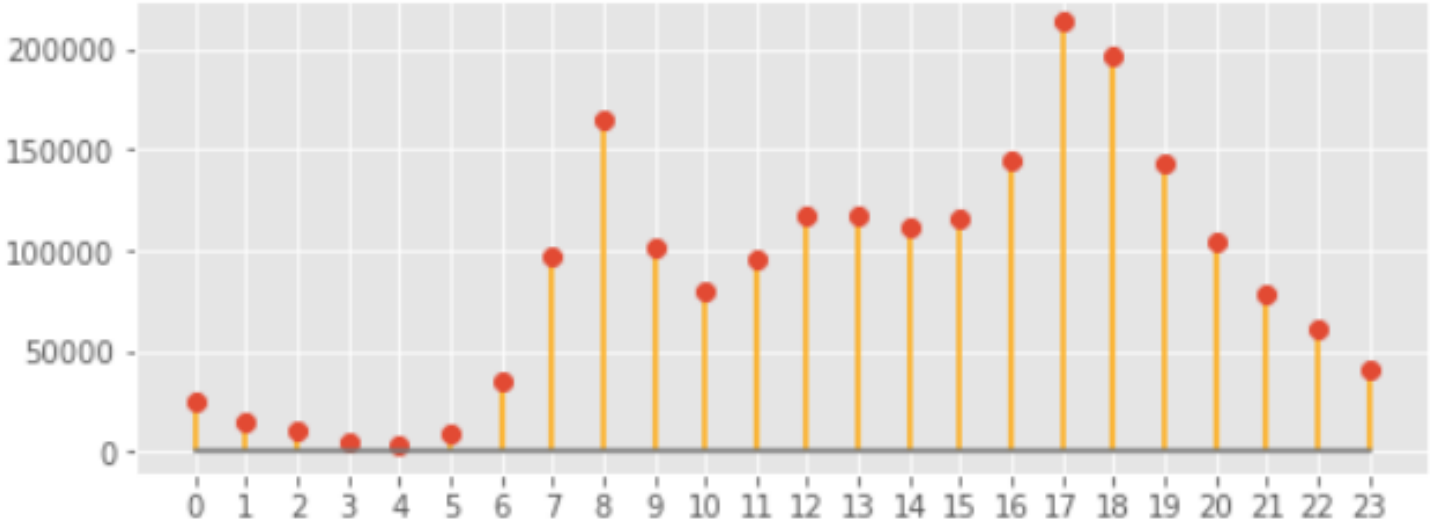
Data Description

Dependent Variable: Hourly count of rental bikes
Data Period: 2011-2012
Data Source: Capital bikeshare system in Washington D.C.

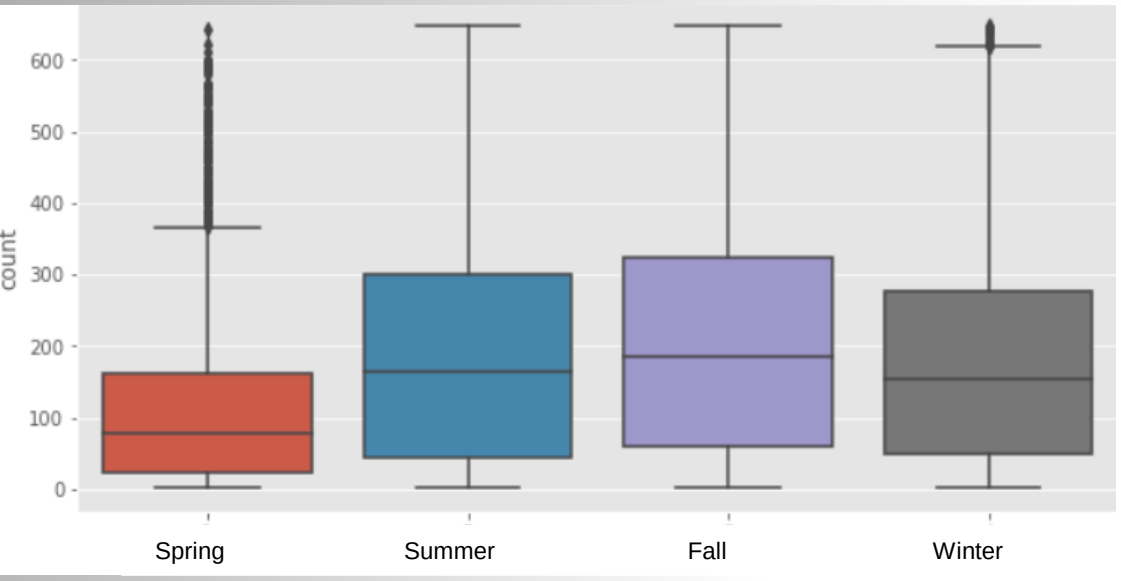
Datetime	: hourly date + timestamp
Season	: 1 = spring, 2 = summer, 3 = fall, 4 = winter
Holiday	: Whether the day is considered a holiday
Workingday	: Whether the day is neither a weekend nor holiday
Weather	: 1: Clear, Few clouds, Partly cloudy, Partly cloudy 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
Temp	: Temperature in Celsius
Atemp	: "feels like" temperature in Celsius
Humidity	: Relative humidity
Windspeed	: Wind speed
Casual	: Number of non-registered user rentals initiated
Registered	: Number of registered user rentals initiated
Count	: Number of total rentals

Data Description

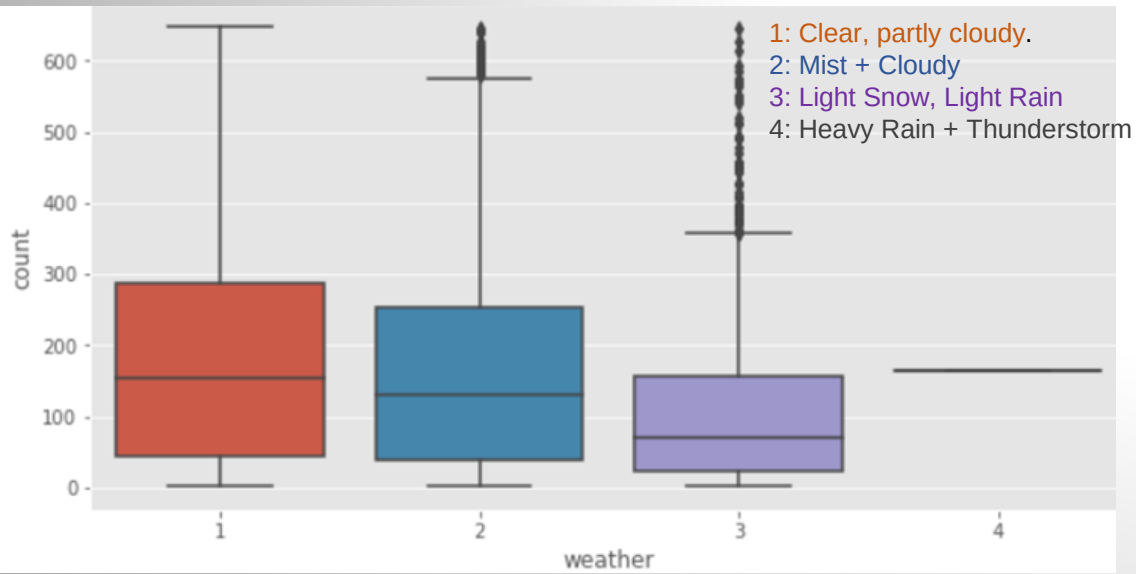




Data Description



!! Spring has outliers, the reason could be changing weather conditions during spring.



In the weather with cloud, mist, light snow, light rain have outliers...

!! The reason for this could be that there could be most likely bad traffic in public transportation in the bad weather…That‘s why people could be prefer to use bikes.

Feature Engineering

	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count	weekday	month	hour
0	1	0	0	1	-1.337036	-1.071073	0.980261	-1.870945	3	13	16	Saturday	January	0
1	1	0	0	1	-1.456365	-1.162906	0.928013	-1.870945	8	32	40	Saturday	January	1
2	1	0	0	1	-1.456365	-1.162906	0.928013	-1.870945	5	27	32	Saturday	January	2

- ✓ Datetime transformed to "weekday" , "month" and "hour".
- ✓ Casual + registered = count, "registered" deleted from data set.
- ✓ "month" deleted from data set, because it is not important variable for modelling.
- ✓ " temp " deleted from data set, because "temp" and "atemp" almost same variable.
- ✓ “Encoding” was applied to the 'weekday’ variable for using it in a model.

	season	holiday	workingday	weather	atemp	humidity	windspeed	casual	hour	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
0	1	0	0	1	-1.071073	0.980261	-1.870945	-0.659419	0	0	1	0	0	0	0
1	1	0	0	1	-1.162906	0.928013	-1.870945	-0.552969	1	0	1	0	0	0	0
2	1	0	0	1	-1.162906	0.928013	-1.870945	-0.616839	2	0	1	0	0	0	0

MODELLING

Independent Variables:

- ✓ Season
- ✓ Holiday
- ✓ Workingday
- ✓ Weather
- ✓ Atemp
- ✓ Humidity
- ✓ Windspeed
- ✓ Causal
- ✓ Hour
- ✓ Weekdays

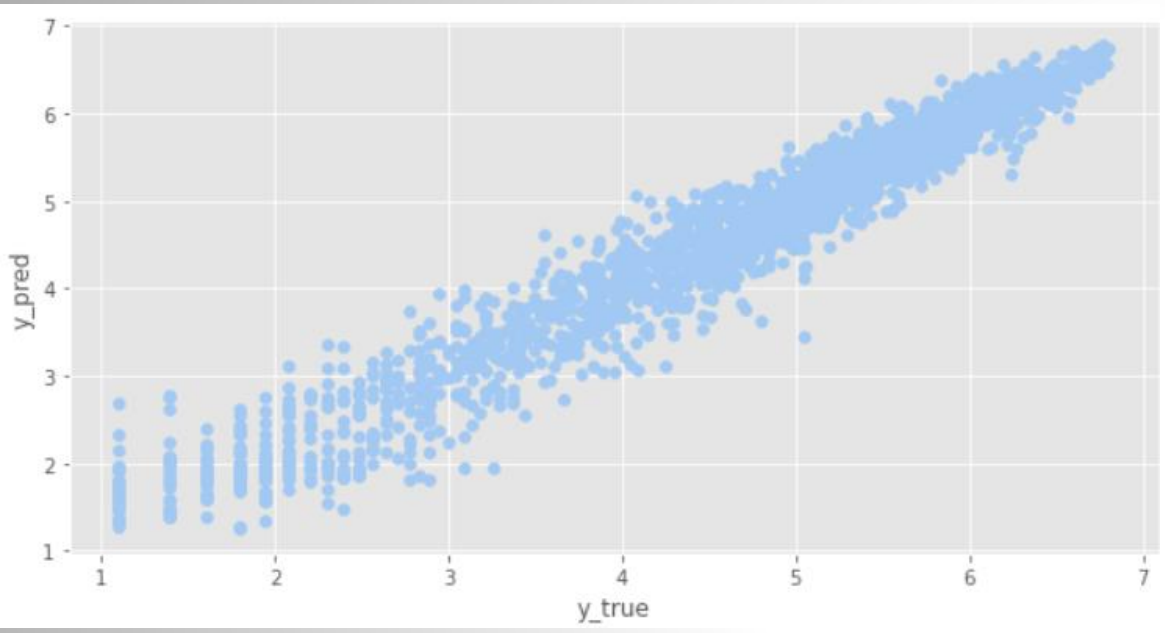
Dependent Variable:

- ✓ Count of rented bikes

Models	**Scores**	**Model 1:** without transfom	**Model 2:** with Power Transform and Standart Scale	**Model 3:** with Power Transform, Sdandart Scale and Logarithmic Transform
Linear Regression	**R2** **RMSE** **MAE**	**0,57** **117,73** **82,48**	**0,63** **92,35** **67,42**	**0,54** **0,90** **0,69**
K Nearest Neighbor Regressor	**R2** **RMSE** **MAE**	**0,66** **104,52** **69,24**	**0,87** **54,92** **37,25**	**0,92** **0,36** **0,27**
Decision Tree Regressor	**R2** **RMSE** **MAE**	**0,89** **58,28** **37,86**	**0,89** **50,21** **33,54**	**0,92** **0,37** **0,27**
Random Forest Regressor	**R2** **RMSE** **MAE**	**0,91** **53,59** **34,19**	**0,90** **47,46** **30,67**	**0,94** **0,31** **0,23** √ Best Model

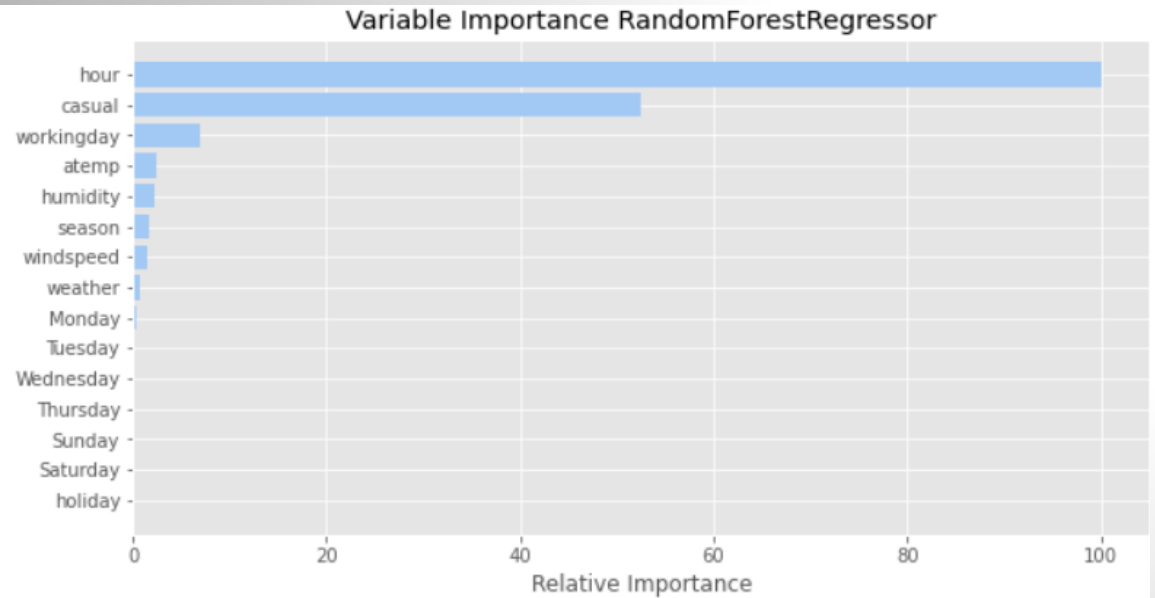
*RMSE: Root Mean Squared Error
*MAE: Mean Absolute Error
*R2: R2 Coefficient of Determination
*All scores were calculeted for the test set.

Random Forest Regressor Model's Performance



Project Goal:
Prediction of the total number of bikes in Washington D.C. that will be rented in each hour.

- ✓ Scatter plot helps us with understanding that **how well match the prediction values and true values**.
- ✓ We can see the reflection of the R2 score of the Random Forest model as high as 91% in this graph.



Hypothesis: The bike sharing are highly related with the time of the day, season and weather conditions.

CONTACT DETAILS

drozkanf@gmail.com

filiz.oezkan@wzb.eu

Please connect to me on Linkedin…