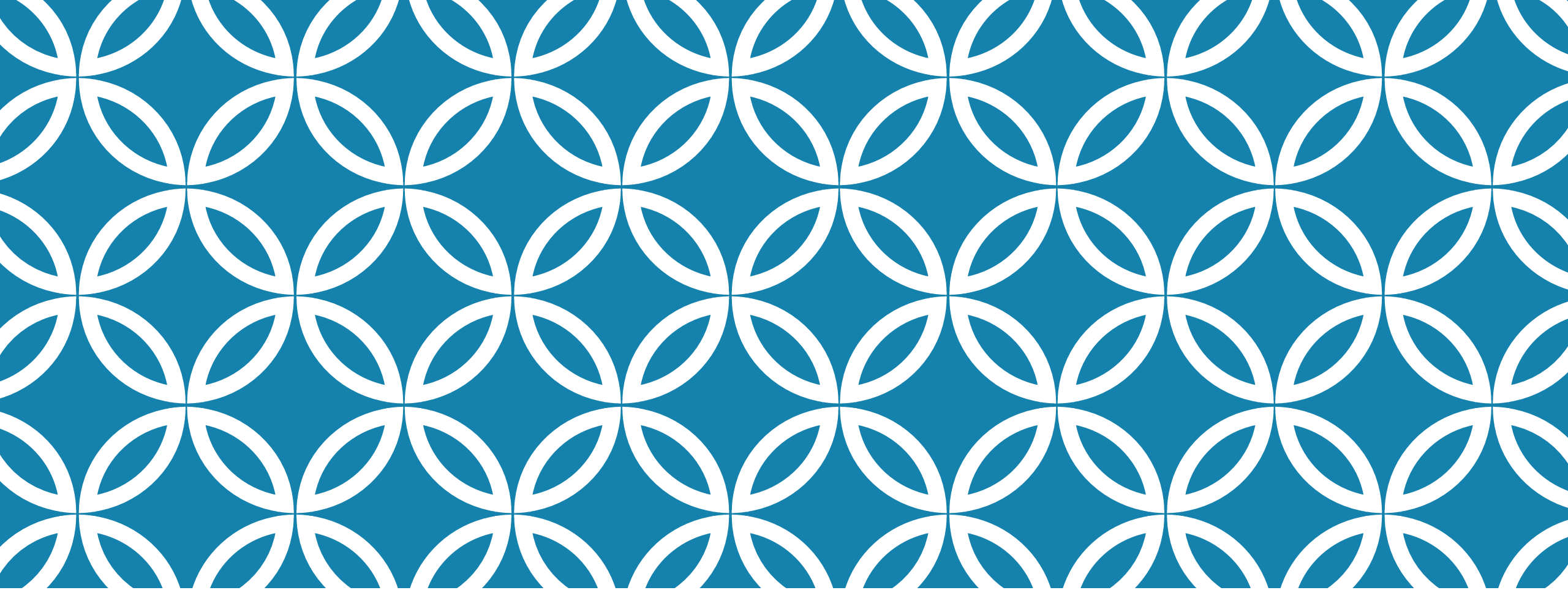


RAINFALL PREDICTION IN AUSTRALIA

Dataset-2007-2017
https://www.kaggle.com/jsphyg/weather-dataset-rattle-package/version/2#weatherAUS.csv

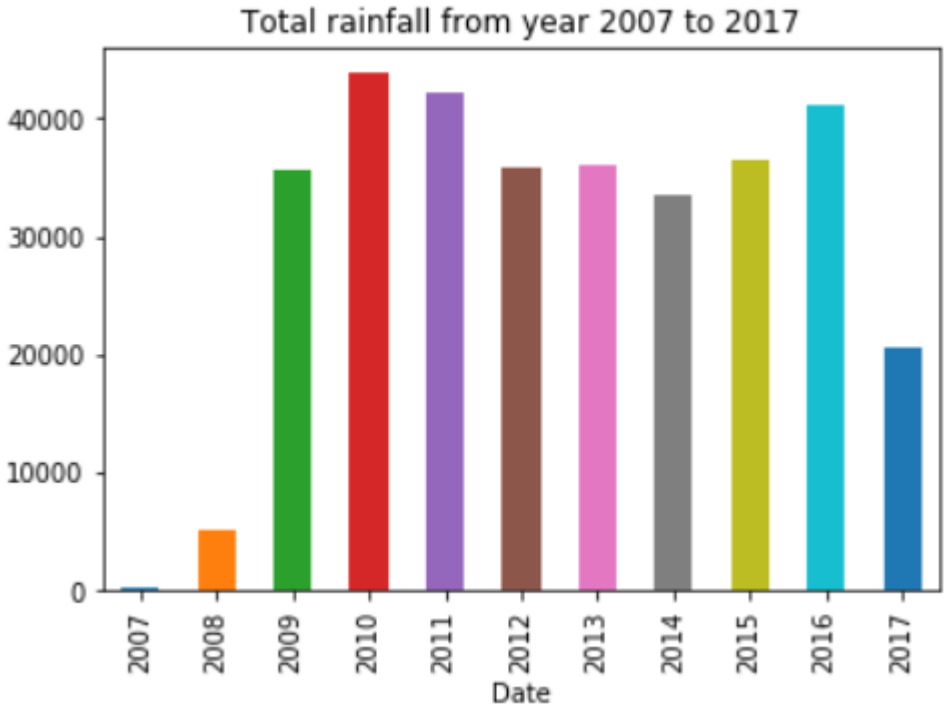
DATA SET EXTRACTION AND CLEAN UP

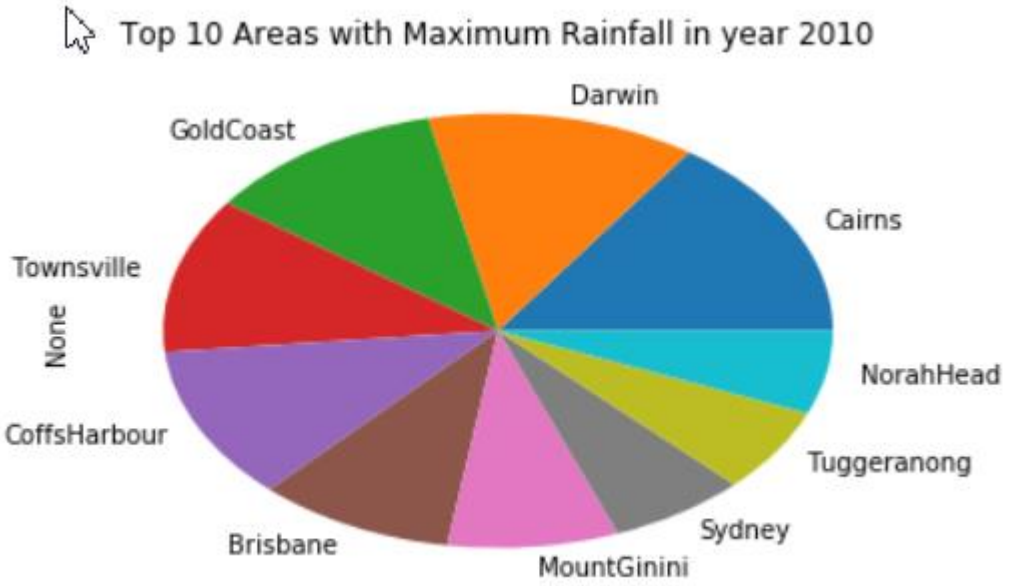
- ✓Dataset is taken from: https://www.kaggle.com/jsphyg/weather-dataset-rattle-package/version/2#weatherAUS.csv
- ✓I have done the Data extraction by downloading and reading the excel file.
- ✓Data has around 150,000 rows and 24 columns.
- ✓Columns Dropped: The columns are not used as I have derived new columns using these.'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm','WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm','Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am','Temp3pm'
- ✓Columns Added Cloud, Pressure, Humidity, WindSpeed
- ✓All the null values 'NaN' is replaced by zeros as for these fields known values can be replaced by the zero for calculations

DESCRIPTIVE STATISTICS

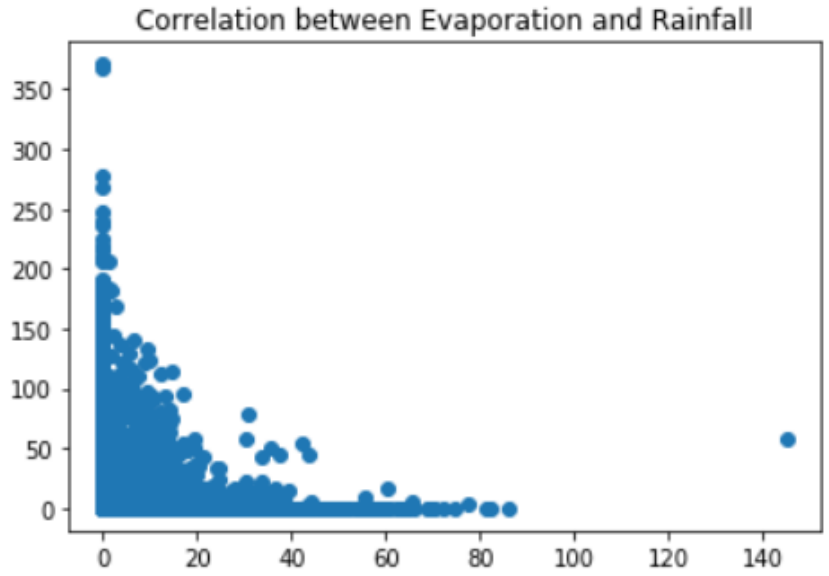
	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	RISK_MM	Pressure	Cloud	Humidity	WindSpeed
count	142193.000000	142193.000000	142193.000000	142193.000000	142193.000000	142193.000000	142193.000000	142193.000000	142193.000000	142193.000000
mean	12.131807	23.174186	2.326738	3.129340	3.988338	2.360682	914.919928	2.554982	58.298105	15.980639
std	6.440548	7.194768	8.426426	4.166674	4.688665	8.477969	304.864442	2.913920	20.657363	7.973465
min	-8.500000	-4.800000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.500000	17.900000	0.000000	0.000000	0.000000	0.000000	1009.750000	0.000000	47.000000	10.500000
50%	12.000000	22.600000	0.000000	1.600000	0.200000	0.000000	1015.450000	1.000000	60.500000	15.000000
75%	16.800000	28.200000	0.600000	5.400000	8.700000	0.800000	1020.550000	5.500000	72.500000	20.500000
max	33.900000	48.100000	371.000000	145.000000	14.500000	371.000000	1040.050000	8.000000	100.000000	83.000000

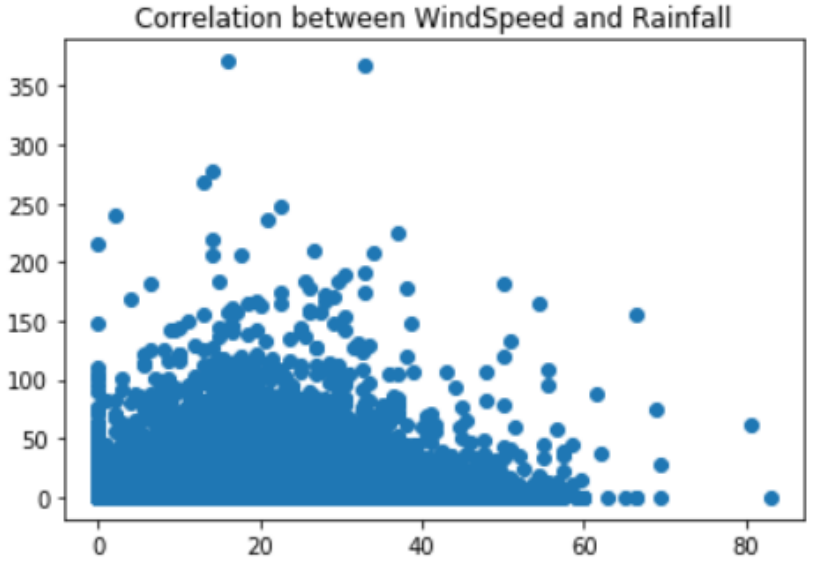
DATA VISUALIZATION CONTINUED...

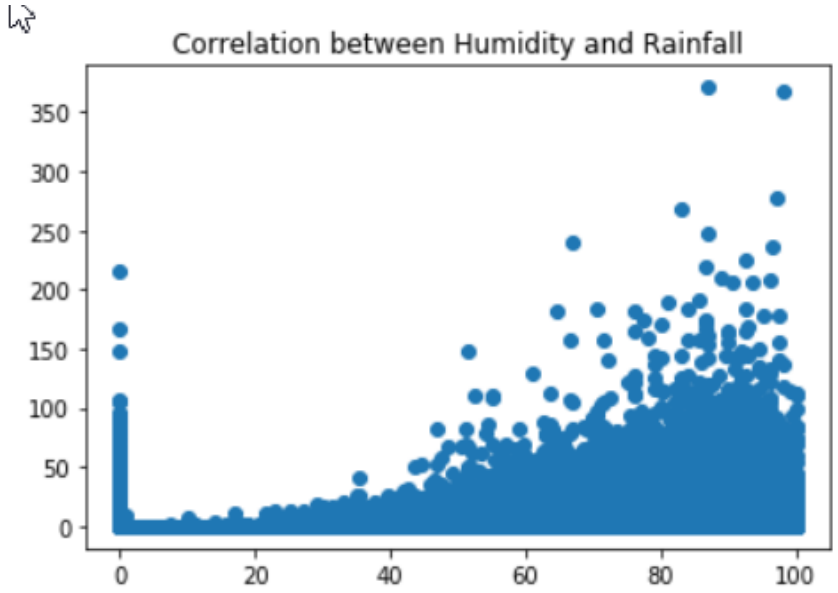




DATA VISUALIZATION CONTINUED...







PREDICTIVE ANALYSIS

```
Intercept   -2.780145
Humidity     0.087599
dtype: float64
```

OLS Regression Results

Dep. Variable:	Rainfall	**R-squared:**	0.046
Model:	OLS	**Adj. R-squared:**	0.046
Method:	Least Squares	**F-statistic:**	6875.
Date:	Tue, 11 Dec 2018	**Prob (F-statistic):**	0.00
Time:	16:33:19	**Log-Likelihood:**	-5.0147e+05
No. Observations:	142193	**AIC:**	1.003e+06
Df Residuals:	142191	**BIC:**	1.003e+06
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2.7801	0.065	-42.545	0.000	-2.908	-2.652
Humidity	0.0876	0.001	82.913	0.000	0.086	0.090

Omnibus:	224358.142	**Durbin-Watson:**	1.467
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	215516986.465
Skew:	10.047	**Prob(JB):**	0.00
Kurtosis:	192.663	**Cond. No.**	185.

```
Intercept    1.049869
WindSpeed    0.079901
dtype: float64
```

OLS Regression Results

Dep. Variable:	Rainfall	**R-squared:**	0.006
Model:	OLS	**Adj. R-squared:**	0.006
Method:	Least Squares	**F-statistic:**	817.5
Date:	Tue, 11 Dec 2018	**Prob (F-statistic):**	2.77e-179
Time:	16:33:48	**Log-Likelihood:**	-5.0442e+05
No. Observations:	142193	**AIC:**	1.009e+06
Df Residuals:	142191	**BIC:**	1.009e+06
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.0499	0.050	21.036	0.000	0.952	1.148
WindSpeed	0.0799	0.003	28.591	0.000	0.074	0.085

Omnibus:	222713.678	**Durbin-Watson:**	1.408
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	198615722.497
Skew:	9.923	**Prob(JB):**	0.00
Kurtosis:	185.015	**Cond. No.**	40.1

```
Intercept      2.601408
Evaporation   -0.101773
dtype: float64
```

OLS Regression Results

Dep. Variable:	Rainfall	**R-squared:**	0.004
Model:	OLS	**Adj. R-squared:**	0.004
Method:	Least Squares	**F-statistic:**	339.3
Date:	Tue, 11 Dec 2018	**Prob (F-statistic):**	1.30e-75
Time:	00:56:43	**Log-Likelihood:**	-2.6798e+05
No. Observations:	81093	**AIC:**	5.360e+05
Df Residuals:	81091	**BIC:**	5.360e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.6014	0.038	68.317	0.000	2.527	2.676
Evaporation	-0.1018	0.006	-18.420	0.000	-0.113	-0.091

Omnibus:	106544.389	**Durbin-Watson:**	1.444
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	29910933.876
Skew:	7.345	**Prob(JB):**	0.00
Kurtosis:	95.933	**Cond. No.**	11.5

CONCLUSION

The rainfall is dependent on the various factors like Evaporation, Humidity, Pressure, Wind and Temperature. And after performing the regression statistics model we can say that rainfall in negatively corelated with Evaporation and Humidity but positively correlated with Windspeed.

THANK YOU