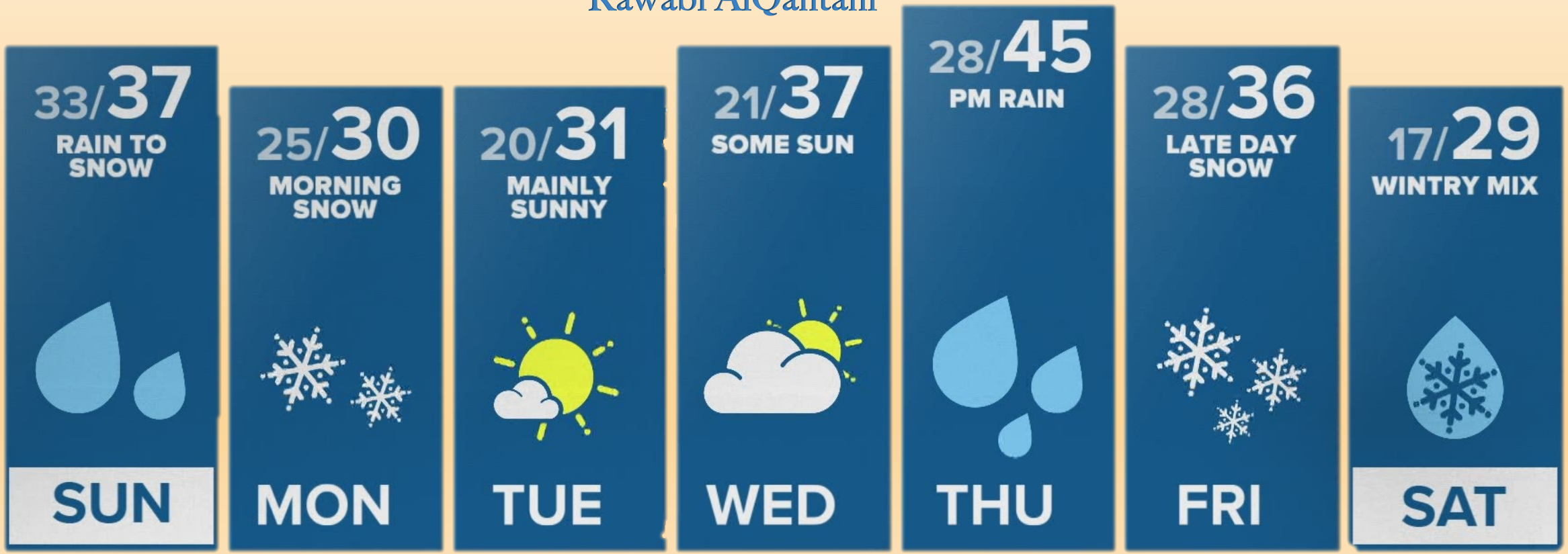
Weather Prediction

Rawabi AlQahtani

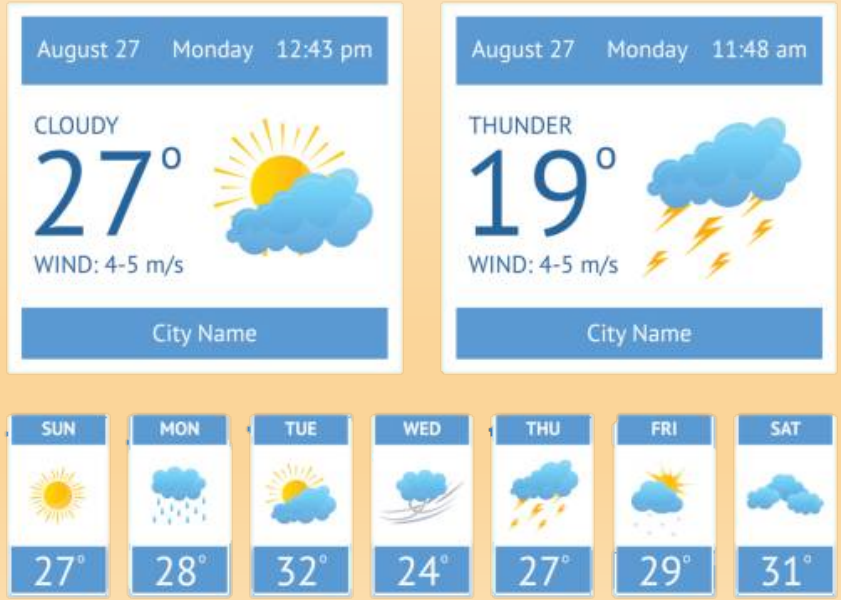


Outline :

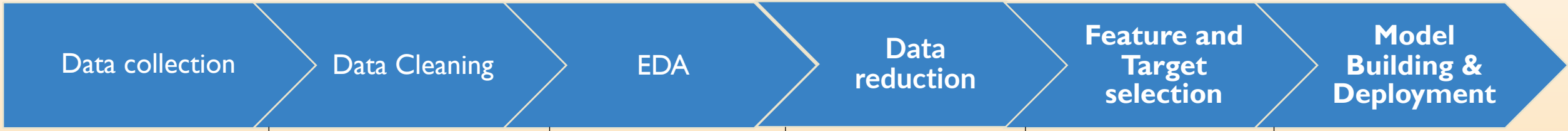
- Project idea
- Work plan
- Time Series
- AutoTS
- PyCaret
- Best model
- Issues
- Future Work
- Q&A
- References

Project Idea :

- Weather prediction for Dammam city, Saudi Arabia.



Work Plan :



Data collection	Data Cleaning	EDA	Data reduction	Feature and Target selection	Model Building & Deployment
Collecting the dataset	Handling the missing Value.	Visualize the data	Dimensionality reduction : reduce the number of input variables in a dataset.	Choosing the target and feature and split the dataset into train and test set	AutoTS PyCaret

Data Collections :

- Dataset : https://www.visualcrossing.com

Date	City	Maximum Temperature	Minimum Temperature	Temperature	Wind Chill	Heat Index

Precipitation	Snow	Snow Depth	Wind Speed	Wind Direction	Wind Gust

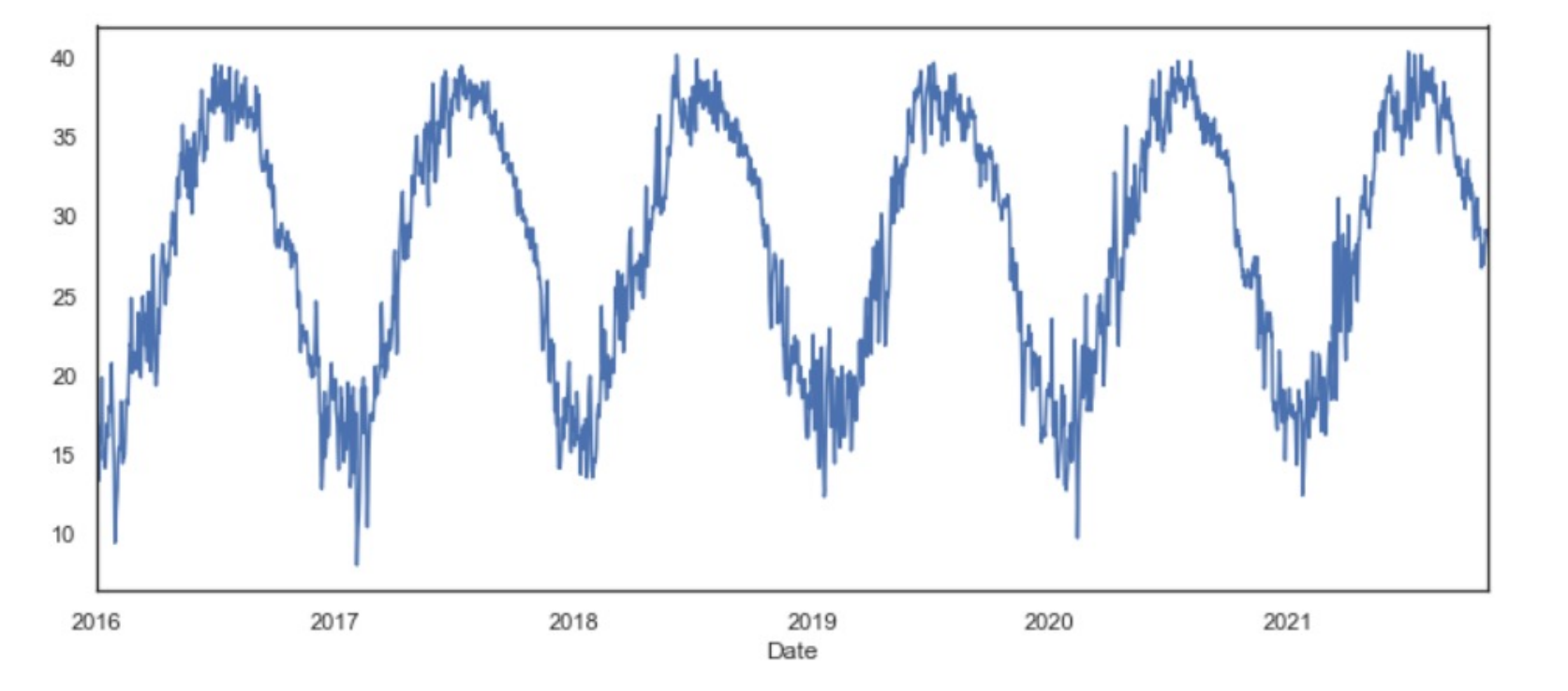
Visibility	Cloud Cover	Relative Humidity	Conditions

Data Cleaning :

- remove messy data
- manage missing values
- remove duplicates and outliers

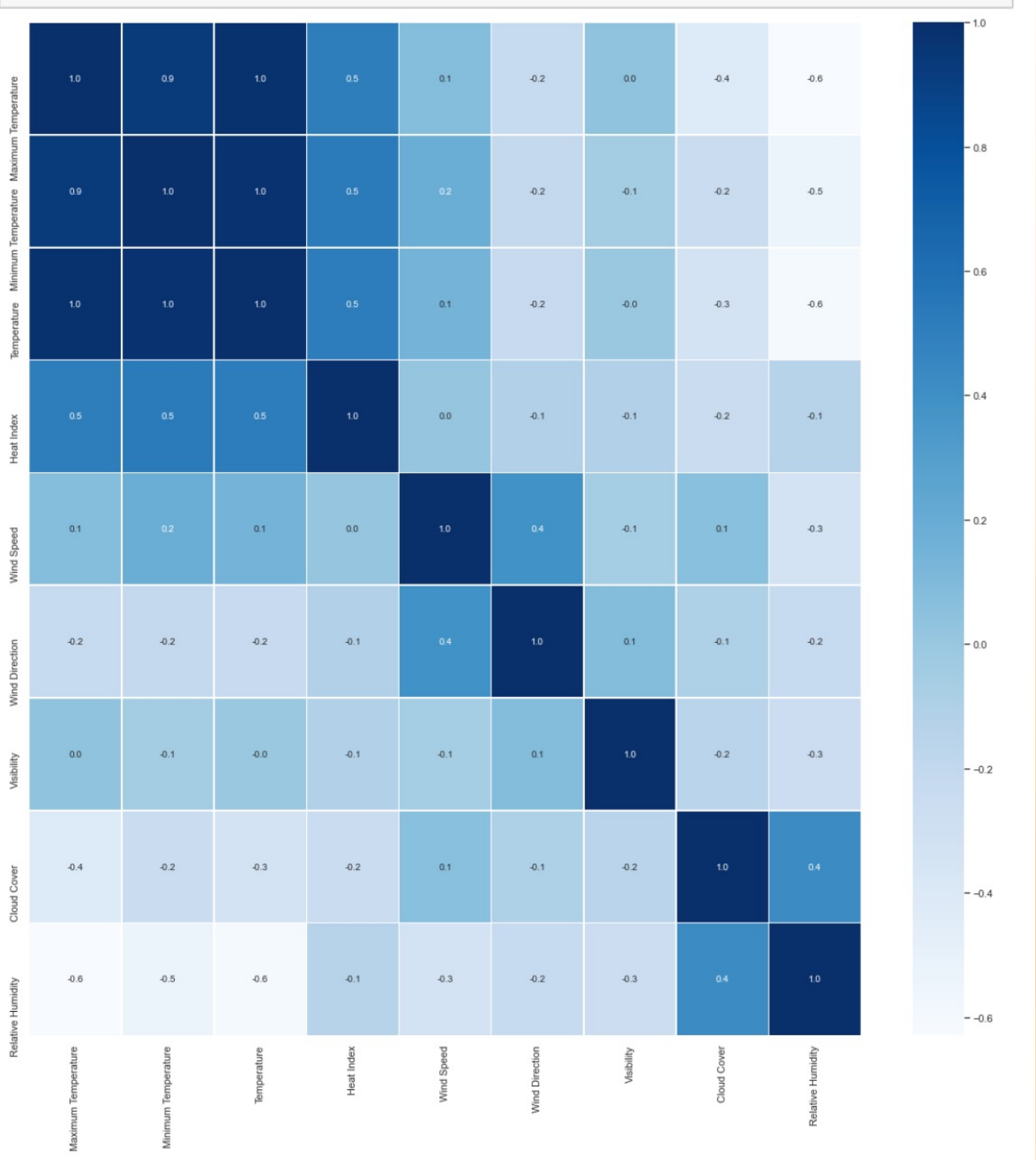
EDA

- Temperature and Date plot



Data Explorations:

- Heatmap



What model is the best model ?

Time Series :

➢ Why time Series ?

- Pattern repeated.

➢ What time Series ?

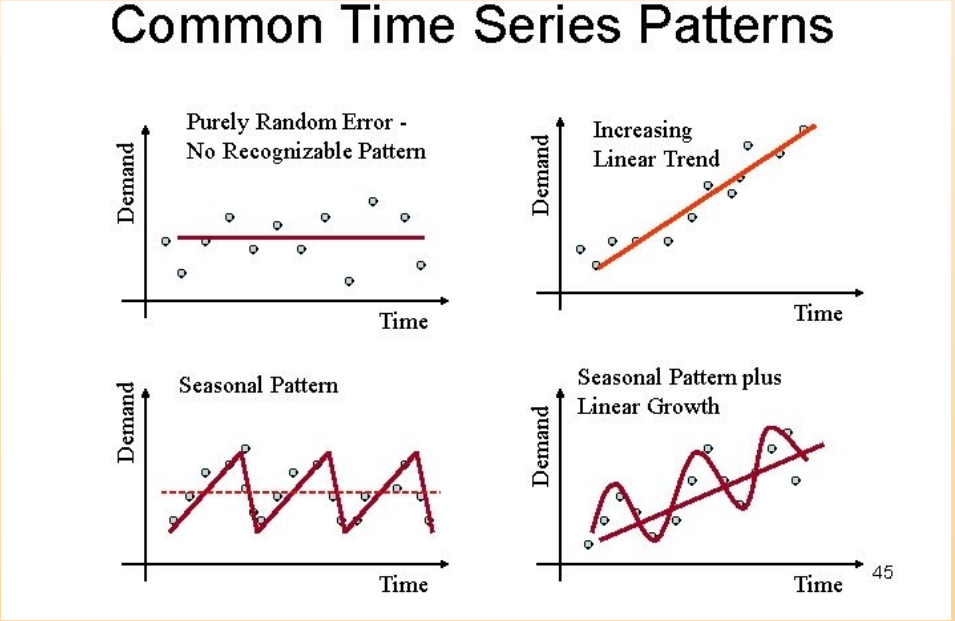
- Time series is a sequence of observations recorded at regular time intervals.
- Time Series Analysis comprised methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.
- Time Series forecasting which is the use of a model to predict future values based on previously observed values.

➢ Importance :

- Understanding past behavior and planning for the future

➢ Components of a Time Series

- ***Trend***
- ***Seasonality***
- ***Cycles***
- ***Irregular fluctuation***



➢ When not to use a time series analysis

- ***When the values are constant.***
- ***Values in the form of a function***

AutoTS Library

➢ Why AutoTs?

➢ What AutoTS?

It is an open-source python library basically used to automate Time Series Forecasting.

➢ Features of Auto-TS library :

- it finds the optimal time series forecasting model
- It trains different models such as ARIMA, SARIMAX, ensemble machine learning models and FB prophet-based models, with all possible hyperparameter configurations, and cross-validation.
- Save time.

```
Leaderboard with best model on top of list:
            name      rmse
2             ML  0.581661
0        Prophet  1.981375
1  auto_SARIMAX       inf
```

Test

Date	yhat
2021-06-01	37.459557
2021-06-02	37.868336
2021-06-03	37.782017
2021-06-04	38.101112
2021-06-05	38.014118
...	...
2021-10-27	27.396517
2021-10-28	27.971649
2021-10-29	27.316912
2021-10-30	28.535494
2021-10-31	28.644131

Forecast

	Temperature
2021-11-01	27.059781
2021-11-02	26.789482
2021-11-03	26.685752
2021-11-04	26.546830
2021-11-05	26.393519
...	...
2022-02-24	19.491261
2022-02-25	19.661698
2022-02-26	19.766187
2022-02-27	19.717202
2022-02-28	19.761517

PyCaret :

➢ Why PyCaret?

➢ What PyCaret

PyCaret is an open-source, low-code machine learning library in Python that automates machine learning workflows.

➢ Features of PyCaret:

- Low-code library.
- Fast and efficient
- Wrap several machine learning libraries and frameworks such as: scikit-learn, XGBoost, LightGBM, CatBoost, Ray, and a few more.

compare_models function trains and evaluates 30+ algorithms from ARIMA to XGboost and more.

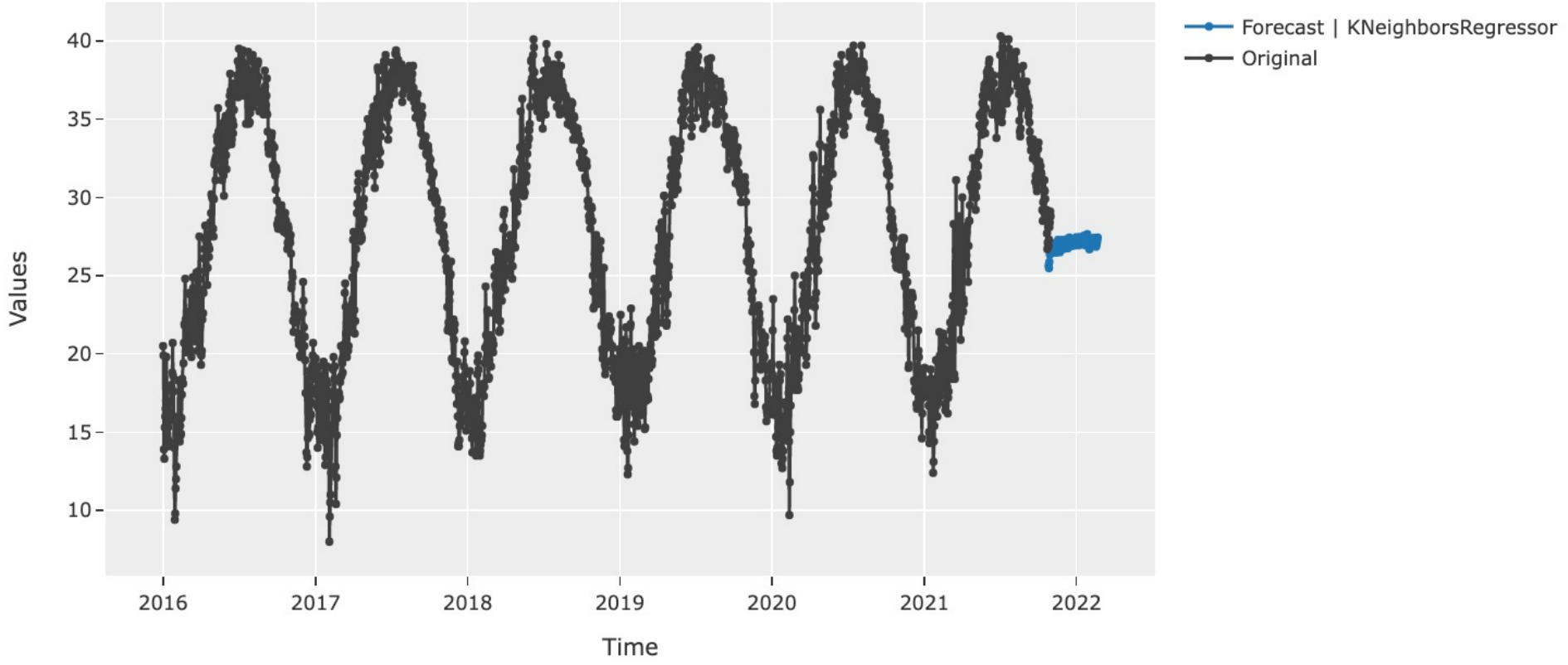
PyCaret :

- After comparing the model. it will use the best model to make the test and the predictions.

	Model	MAE	RMSE	MAPE	SMAPE	R2	TT (Sec)
knn_cds_dt	K Neighbors w/ Cond. Deseasonalize & Detrending	0.8290	1.0016	0.0286	0.0280	-0.7753	0.0533
xgboost_cds_dt	Extreme Gradient Boosting w/ Cond. Deseasonalize & Detrending	1.2196	1.4305	0.0422	0.0410	-2.3161	0.5767
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize & Detrending	1.2228	1.4194	0.0425	0.0411	-2.0787	0.5733
et_cds_dt	Extra Trees w/ Cond. Deseasonalize & Detrending	1.2252	1.4329	0.0421	0.0412	-2.9778	0.1967
rf_cds_dt	Random Forest w/ Cond. Deseasonalize & Detrending	1.2312	1.4531	0.0425	0.0414	-2.5234	0.3167
lasso_cds_dt	Lasso w/ Cond. Deseasonalize & Detrending	1.2730	1.4901	0.0443	0.0428	-2.3231	0.0300
auto_arima	Auto ARIMA	1.2743	1.5249	0.0441	0.0428	-2.7926	52.1333
omp_cds_dt	Orthogonal Matching Pursuit w/ Cond. Deseasonalize & Detrending	1.3054	1.5100	0.0455	0.0439	-2.3823	0.0267
theta	Theta Forecaster	1.3142	1.5289	0.0458	0.0442	-2.4600	0.0600
en_cds_dt	Elastic Net w/ Cond. Deseasonalize & Detrending	1.3200	1.5204	0.0459	0.0443	-2.4753	0.0300
naive	Naive Forecaster	1.3238	1.5191	0.0462	0.0445	-2.3995	0.0133
lr_cds_dt	Linear w/ Cond. Deseasonalize & Detrending	1.3255	1.5824	0.0460	0.0445	-3.0008	0.0267
ridge_cds_dt	Ridge w/ Cond. Deseasonalize & Detrending	1.3255	1.5823	0.0460	0.0445	-3.0000	0.0267
br_cds_dt	Bayesian Ridge w/ Cond. Deseasonalize & Detrending	1.3257	1.5816	0.0460	0.0445	-2.9914	0.0267
lar_cds_dt	Least Angular Regressor w/ Cond. Deseasonalize & Detrending	1.3262	1.5850	0.0460	0.0445	-3.0224	0.0400
exp_smooth	Exponential Smoothing	1.3398	1.5459	0.0468	0.0450	-2.5090	0.2100
ets	ETS	1.3397	1.5458	0.0468	0.0450	-2.5080	0.1833
gbr_cds_dt	Gradient Boosting w/ Cond. Deseasonalize & Detrending	1.3650	1.5654	0.0475	0.0458	-2.6584	0.2433
llar_cds_dt	Lasso Least Angular Regressor w/ Cond. Deseasonalize & Detrending	1.3967	1.5532	0.0468	0.0469	-7.8857	0.0267
polytrend	Polynomial Trend Forecaster	1.4032	1.5582	0.0470	0.0471	-8.1916	0.0100
huber_cds_dt	Huber w/ Cond. Deseasonalize & Detrending	1.4347	1.6695	0.0497	0.0480	-3.3516	0.0433
ada_cds_dt	AdaBoost w/ Cond. Deseasonalize & Detrending	1.4633	1.6655	0.0508	0.0490	-3.3477	0.1233
snaive	Seasonal Naive Forecaster	1.5571	1.9411	0.0537	0.0516	-5.7614	0.0167
arima	ARIMA	1.6360	1.9891	0.0561	0.0543	-6.7084	0.1133
grand_means	Grand Means Forecaster	2.1250	2.2323	0.0692	0.0723	-25.7169	0.0133
dt_cds_dt	Decision Tree w/ Cond. Deseasonalize & Detrending	2.1936	2.4497	0.0738	0.0741	-22.8405	0.0333
par_cds_dt	Passive Aggressive w/ Cond. Deseasonalize & Detrending	2.4624	2.7991	0.0825	0.0836	-24.3785	0.0267

Best model :

- According to MAE is KNN



Forecast

2021-11-01	28.8047
2021-11-02	28.2070
2021-11-03	28.4894
2021-11-04	28.6512
2021-11-05	28.4924
	...
2022-02-24	28.1176
2022-02-25	28.3764
2022-02-26	28.9585
2022-02-27	28.1221
2022-02-28	28.3809

Future work

- Condition
- Cover more cities

Issues :

- Dataset
- Model
- VS Code.


مسك المهارات
Misk Skills


وزارة الاتصالات
وتقنية المعلومات
MINISTRY OF COMMUNICATIONS
AND INFORMATION TECHNOLOGY


Q&A

References :

- https://www.visualcrossing.com/weather/weather-data-services#/editDataDefinition
- https://facebook.github.io/prophet/docs/quick_start.html
- https://www.analyticsvidhya.com/blog/2021/04/automate-time-series-forecasting-using-auto-ts/
- https://towardsdatascience.com/announcing-pycarets-new-time-series-module-b6e724d4636c