2022

جامعة جدة
University of Jeddah

Final Senior Project Report

DONE BY:
AHMED ALMDAIFER
ABDULAZIZ MGARRY
RAAD SABBAGH

SUPERVISOR:
DR. OMAR AL-GHUSHAIRY

TABLE OF CONTENT

1.1 Introduction

High wind causes many significant negative impacts on society, the economy, and the environment on a local, regional, and global scale. Strong winds can generate many factors like sand and dust storms and sea hurricanes. The environmental and health risks of such storms cannot be permanently reduced, but their impact can be reduced by taking appropriate precautions.

1.2 Problem Definition

High winds can cause trees to fall, debris to fly, and buildings to collapse, which can lead to power outages, transportation disruptions, damage to buildings and vehicles, and injury or death. It also can cause dusty waves, often in the afternoon and evening hours, with which the horizontal visibility significantly decreases specifically Kingdom of Saudi Arabia.

Health effects:

- People with respiratory system difficulties
- Elderly people
- Heart-diseased people
- Pregnant women
- People who work outdoor
- Road accidents and aviation risks due to poor visibility

Economic Effects:

- The cost of restoring damaged buildings.
- Costs associated with removing sand and dust burial during storms for infrastructure works such as oil pipelines.
- The cost of repairing damaged electrical currents.
- Ship ports destroyed by sea hurricanes.
- flying debris and building collapses.
- Delayed takeoff and landing of aircraft, and delays may occur in the marine shipping route.

1.3 Proposed Solution (The Recommended Solution)

The main motive for doing this project is to forecast the wind speed and direction one day ahead (24h). We selected the west-cost of the Saudi Arabia area as a station for our analysis. With the tools being used are Machine Learning methods andalgorithms and Microsoft Excel for the datasets.

1.4.1 Aims

The project aims to forecast the wind speed and direction several hours to one week ahead, looking if the wind speed affects the visibility, temperature, and Humidity. By using machine learning algorithms.

1.4.2 Objectives

We will be dealing with this situation by:

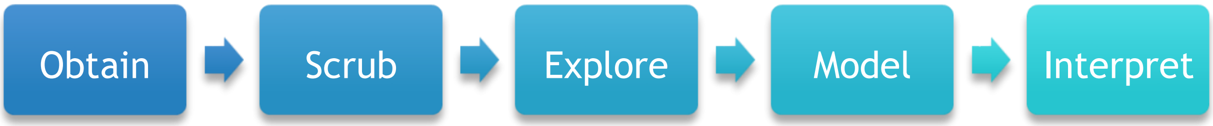
1- Collecting data from reliable sources.
2- Forecasting wind speed.
3- Forecasting wind direction.
4- Using machine learning algorithms and comparison with previous works.
5- Measure sure if wind speed can affect vision or not

1.5 Target users

We **think that this project will have many beneficiaries for example:**

1- Citizen
2- Governmental and private establishments.
3- National Meteorological Center.
4- Airports and ports

1.6 Project Methodology


Obtain
Scrub
Explore
Model
Interpret

1- Obtain the data and gather it from relevant sources.
2- Clean the data to formats that the machine understands.
3- Find significant patterns and trends using statistical methods.
4- Construct models to predict and forecast.
5- Put the results into good use.

1.7 Project Plan

We will locate every station in the west of the kingdom of Saudi Arabi and analyze each one to figure out which station record the highest wind speed, by using the tools we learned through the time our study in the university to reduce the impact of future high wind speed

1.8 Project Timeline

			PHASE ONE			PHASE TWO			PHASE THREE			PHASE FOUR			FINAL PHASE		
Chapter	**TASK**	**% DONE**	**WEEK 1**	**WEEK 2**	**WEEK 3**	**WEEK 4**	**WEEK 5**	**WEEK 6**	**WEEK 7**	**WEEK 8**	**WEEK 9**	**WEEK 10**	**WEEK 11**	**WEEK 12**	**WEEK 13**	**WEEK 14**	**WEEK 14**
1	Project Conception and Initiation		◉														
1.1	Choose team members	100%		◉	◉												
1.1.1	Choose a supervisor and idea	100%		◉	◉												
1.2	Chapter 1 Submission	100%				◉	◉										
2	Project Definition and Planning																
2.1	Review the State of the Arts	5%						◉	◉								
2.2	Literature review	100%						◉	◉								
2.3	Chapter 2 Submission	0%						◉	◉								
3	Project Launch and Execution																
3.1	Data Collection	0%		◉	◉	◉	◉	◉	◉	◉	◉						
3.2	Data Description	0%				◉	◉	◉	◉	◉	◉						
3.2.1	Data Exploration	0%					◉	◉	◉	◉	◉						
4	Project Performance / Monitoring																
4.1	Removal of Unwanted Observations	0%										◉	◉				
4.2	Handling Missing Data	0%										◉	◉				
4.3	Data Transformation	0%										◉	◉				
Results and discussions													◉	◉			
Final report submission														◉			
presentations															◉	◉	

1.9 Tools and requirements

We will be using the best tools for our project to get the best results as data scientists. These tools are as follows:

- Machine learning:
    - Python:
        - Random Forest
        - SVM
        - Pandas
        - Logistic Regression
        - NumPy
- MS Excel:
    - Dataset

1.10 Conclusion

At the end of our project, we hope that we give real solutions from what we taught all over the years at the university. High wind speed is a real danger and should be fought with science, while using the tools we learned from our journey, to help at least reduce the impact of the High wind speed.

2022

جامعة جدة

University of Jeddah

Chapter #2

2.1 Introduction

The purpose of this chapter is to provide background information about forecasting wind speed, wind direction, and many other tools and techniques. As part of the comparison of our project and the related work, we will include related work from similar projects that have been undertaken.

2.2 Background

Random Forest:

A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and used for Classification and Regression problems in Machine Learning.

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T}$$

* R, S.E. (2022) *Random Forest: Introduction to random forest algorithm, Analytics Vidhya*. Sruthi E R. Available at: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/ (Accessed: November 17, 2022).

Support vector machine:

A support vector machine (SVM) is a machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that looksat data and sorts it into one of two categories.

$$W^T.x = 0$$

* Stecanella, B. (2017) *Support Vector Machines (SVM) algorithm explained, MonkeyLearn Blog*. Bruno Stecanella. Available at: https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/ (Accessed: November 17, 2022).

Logistic Regression:

The logistic regression model is used to predict the probability of the occurrence of an event and the adequacy of data on the logistic curve. Logistic regression has severalexpected variables that can have digital or factional use.

ln =0+ß1 1+...+ßn

1 –

* Lawton, G., Burns, E. and Rosencrance, L. (2022) *What is logistic regression? - definition from Searchbusinessanalytics*, *SearchBusinessAnalytics*. TechTarget. Available at: https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression (Accessed: November 17, 2022).

Pandas Library:

is a software library written for the Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series.

* Resources, M. (2016) *Pandas: Python library - mode*, *Mode Resources*. Mode Resources. Available at: https://mode.com/python-tutorial/libraries/pandas/ (Accessed: November 17, 2022).

NumPy Library:

NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices.NumPy was created in 2005 by Travis Oliphant. It is an open-source project and you can use itfreely.NumPy stands for Numerical Python.

*w3schools (no date) *Numpy introduction*, *Introduction to NumPy*. w3schools. Available at: https://www.w3schools.com/python/numpy/numpy_intro.asp (Accessed: November 17, 2022).

2.3 Related work

The main idea of our project is to focus on forecasting wind speed, wind direction, andatmosphere. It is a technique for forecasting high wind speed and a sort of predictinganalysis that identifies wind speed before 24h Approximately, to one week in Python. This tool and many others will be used for data preprocessing and machine learning algorithms.

As we will be using three models in machine learning Random Forest will be used for wind direction, a Support vector machine will be used for wind speed, and Logistic Regression will be used in both wind speed and wind direction to get the best accuracy for both models.

wind power or energy, combining the literature can be divided into four categories: Ultra-short-term forecast: From a few minutes to one hour ahead, Short-term forecast: From one hour to several hours ahead, medium-term forecast: From several hours to one week ahead and Long-term forecast: From one week to one year or more ahead.

First Literature:

We found a similar project and they used a short-term forecast as we will be using a midterm forecast. We shared some methods to reach our goal such as Random Forest used for the wind direction and Logistic regression for the wind speed in both projects. These algorithms gave the best results and have been used for all analyzed cases. As for the dataset, we collected our data from King Abdullah Petroleum and Research Center, they collected their dataset from World Atlas.

Our primary goal is to forecast wind speed and direction in the west of the Kingdom of Saudi Arabia while using 3 methods, to avoid any future disasters. Additionally, our project aims on forecasting the wind speed and wind direction several hours to one week ahead, looking if the wind speed affects the visibility, temperature, and Humidity. The other project, their goal is to prove the competency of a minimalistic approach to wind prediction. Using only two attributes decreases the amount of data, provides results faster, and can be very indication for wind power plant control.

Second Literature:

The goal of the second similar project was environmental considerations that have prompted the use of wind power as a renewable energy resource. The biggest challenge in integrating wind power into the electric grid is its intermittency.

taking several time-scale classifications with different methods mainly based on Artificial Neural Networks (ANN).

The methods used in the literature are:

- **The persistent Method**, this method is also known as Naïve Predictor. it is more accuratethan most of the physical and statistical methods for very short to short-term forecasts.
- **Physical Approach**, physical systems use parameterizations based on a detailed physical description of the atmosphere. Numeric Weather Prediction (NWP): This method is a physical approach to wind forecasting. NWP models operate by solving complex mathematical models that use weather data like temperature, pressure, surface roughness, and obstacles. And it is used for medium to long-term forecasts.
- **Statistical Approach**, the statistical approach is based on training with measurement data and uses the difference between the predicted and the actual wind speeds in the immediate past to tune model parameters. Sub-classification of this approach is: Time-series based models, and neural network (NN) based methods. Auto-Regressive Moving Average (ARMA) models are the most popular type in the time-series based approach to predicting future values of wind speed or power.
- **The hybrid Approach**, in general, combination of different approaches such as mixing physical and statistical approaches or combining short-term and medium-term models,etc., is referred to as a hybrid approach.

Third Literature:

The goal for the third is to explore and evaluate how a selection of supervised ML models, that vary in complexity and design, can be used for wind power production forecasting 1-48 h ahead using meteorological NWP forecast data of several wind features and historical power output data from multiple wind parks in the same geographical region.

4th Literature:

Digital design and analysis tools are continually progressing, enabling more seamless integration of climatic impacts into the conceptual design stage, which naturally means the enhanced environmental performance of the final designs. Planning sustainable urban configurations and, consequently, environment-derived architectural forms become more rapid and requires less effort enabling smooth incorporation into day-to-day practice. This research paper presents a wind prediction-based architectural design method for improving outdoor wind comfort through urbanism and architecture. The added value of the environment-driven design loop consisting of parametric design, wind flow analysis, and necessary design modifications lies in leveraging the newly developed wind prediction tool InFraRed. As is demonstrated in the application study in Kosice, Slovakia, iterating through various design options and evaluating their impact on the wind flow is swift and reliable. That enables the designer to explore the best-performing design alternatives for outdoor wind comfort, yet the extra time required for the analysis is negligible.

5th Literature:

Emanating from the base of the Sun's corona, the solar wind fills the interplanetary medium with a magnetized stream of charged particles whose interaction with the Earth's magnetosphere has space weather consequences such as geomagnetic storms. Accurately predicting the solar wind through measurements of the spatiotemporally evolving conditions in the solar atmosphere is important but remains an unsolved problem in heliophysics and space weather research. In this work, we use deep learning for the prediction of solar wind (SW) properties. We use extreme ultraviolet images of the solar corona from space-based observations to predict the SW speed from the National Aeronautics and Space Administration (NASA) OMNIWEB data set, measured at LagrangianPoint 1. We evaluate our model against autoregressive and naive models and find that our model outperforms the benchmark models, obtaining a best-fit correlation of 0.55 ±
0.03 with the observed data. Upon visualization and investigation of how the model uses data to make predictions, we find higher activation at the coronal holes for fast wind prediction (≈3 to 4 days prior to prediction), and at the active regions for slow wind prediction. These trends bear an uncanny similarity to the influence of regions potentially being the sources of fast and slow wind, as reported in the literature. This suggests that our model was able to learn some of the salient associations between coronal and solar wind structures without built-in physics knowledge. Such an approach may help us discover hitherto unknown relationships in heliophysics data sets.

6th Literature:

Solar wind modeling is categorized into empirical and physics-based models that both predict the properties of solar wind in different parts of the heliosphere. Empirical models are relatively inexpensive to run and have shown great success at predicting the solar wind at the L1 Lagrange point. Physics-based models provide more sophisticated scientific modeling based on magnetohydrodynamics (MHD) that are computationally expensive to run. In this paper, we propose to combine empirical and physics-based models by developing a physics-guided neural network for solar wind prediction. To the best of our knowledge, this is the first attempt to forecast solar wind by combining data-driven methods with physics constraints. Our results show the superiority of our physics-constrained model compared to other state-of-the-art deep-learning predictive models.

7th Literature:

Wind prediction is a key factor for improving the estimated time of arrival (ETA)accuracy at a specific waypoint. The authors propose a new wind prediction method using numerical weather forecasts and past SSR Mode-S wind information of nearby aircraft. The spatial component is interpolated, and the temporal component is extrapolated by Gaussian Process Regression. The proposed method estimates the wind error from the numerical weather forecast instead of the wind magnitude, which allows the wind prediction accuracy tobe up to the numerical weather forecast. The result shows that the wind prediction accuracy has been improved by the proposed method, and the corresponding ETA accuracy has been improved by about 15% compared to the method using numerical weather forecasts only.

8th Literature:

Nowadays, wind power is playing a significant role in power systems; it is necessary to improve prediction accuracy, which will help make better use of wind sources. The existing neural network methods, such as recurrent neural networks (RNN), have been widely used in wind prediction; however, RNN models only consider the dynamic change of temporal conditions and ignore the spatial correlation. In this work, we combine the graph convolutional neural (GCN) with the gated recurrent unit (GRU) to do predictions on simulated and real wind speed and wind power data sets.

9th Literature:

The application of deep learning to wind time series for multi-step prediction obtains good results at short horizons. The accuracy of a wind forecast is highly dependent on the specific structure of wind in the specific location, as many local features influence wind behavior. The characterization of the complexity of a site for wind prediction is defined as forecastability or predictability and can be obtained from the inner structure of the meteorological time series observations from a site. We analyze the time series structure searching for properties that have a high correlation with the prediction result, and properties that can create measures that have the potential to describe the forecastability of a site. The best measures will show a high correlation with the accuracy of the predictions. In this work, we analyze wind time series from 126,692 wind locations in the US, where we apply several deep learning methods first, and then we verify several forecastability descriptors with the accuracy of deep learning results. We require High-Performance Computing (HPC) resources for this task as the deep learning algorithms have sensible resource requirements and are applied to a large set of data. The measures defined and explored in this work are based on several techniques that decompose or transform the wind time series. By combining several of these measures, we can obtain better predictors of the site complexity, which will allow us to evaluate the future error of a prediction on this site. Forecastability measures can contribute to a wind site's multi-dimensional description, becoming a valuable tool for wind resource analysts and wind forecasters.

10th Literature:

The wind speed prediction in Kudat, Malaysia had been done by using the Mycielski-1 approach and K-mean clustering statistical method. There is some improvement in obtaining the random number of Mycielski-1. Besides, the comparison of K-means clustering with the optimal number of K is presented in this paper. Wind prediction is important to study a favorable site's wind potential. The prediction is based on 3 yearsof historical data provided by the Meteorology Department of Malaysia and 1-year data as the reference to check the accuracy of both algorithms. The basic concept of the Mycielski-1 algorithm is to predict the next value by looking at historical data. Meanwhile, the K-meansclustering can group the values with similar means into the same group, and the prediction can be done by getting the probability of occurrence. The result shows the prediction of the Mycielski-1 algorithm and K-means clustering are promising. The wind speed is predicted to obtain the mean power for energy planning.

2.4 Comparison between the proposed system and the literature

First Comparison:

	Tools	Methodology	End Goal
Our project	• Python • Random Forest • Support vector machine • Logistic Regression	• Obtain the data and gather it from relevant sources. • Clean the data to formats that the machineunderstands. • Find significant patterns and trends using statistical methods. • Construct models to predict and forecast. • Put the results into good use.	Forecasting the wind speed and direction in the west of Saudi Arabia to avoid future disasters.
Similar project#1	• Support vector machine • Logistic Regression • MP neural network • Random Forest	• Time periods were chosen to show the difference in results for different time periods. • analyze different time intervals. • Collecting data and training models	To prove the competencyof the minimalistic approach to wind prediction sing only two attributes.

Second Comparison:

	Tools	Methodology	End Goal
Our project	• Python • Random Forest • Support vector machine • Logistic Regression	• Obtain the data and gather it from relevant sources. • Clean the data to formats that the machineunderstands. • Find significant patterns and trends using statistical methods. • Construct models to predict and forecast. • Put the results into good use.	Forecasting the wind speed and direction in the west of Saudi Arabia to avoid future disasters.
Similar project#2	• Persistent Method • Physical Approach • Statistical Approach • Hybrid Approach	• Naïve Predictor for very short to short-term forecasts. • NWP models operate by solving complex mathematical models. • ARMA predicts futurevalues of wind speed or power. • Combination of different approaches combining short term and medium-term models	Environmental considerations have prompted the use of wind power asa renewable energy resource.

Third Comparison:

	Tools	Methodology	End Goal
Our project	- Python - Random Forest - Support vector machine - Logistic Regression	- Obtain the data and gather it from relevant sources. - Clean the data to formats that the machineunderstands. - Find significant patterns and trends using statistical methods. - Construct models to predict and forecast. - Put the results into good use.	Forecasting the wind speed and direction in the west of Saudi Arabi to avoid future disasters
Similar project#3	- Time Series Forecasting. - Linear Regression. - Linear Regression with LASSO. - Autoregressive model (AR). - Bagging and Random Forest Regression (RFR).	- Data pre-processing - LASSO, RFR GB. - Training- and Test data. - Cross-validation with time series split. - Hyperparameter tuning.	explore and evaluate how a selection of supervised ML models, that vary in complexity and design, can be used for wind power production.

2.5 Conclusion

The comparisons were fruitful since they helped give us ideas about various tools and methods, making the road ahead much clearer, and therefore our work and analysis willbe improved.

Although our project's main goal is different from the similar kinds of literature, the process isvery similar especially when we apply the machine learning methods, it will help us reach our project's goal, and we will add more methods and tools as we proceed.

2.6 Literature’s References

- First Literature:

  Music, E., Halilovic, Jusufovic, Kevric (2018) *Wind direction and speed prediction using machine learning*, *Research Gate*. International BURCH University. Available at: https://www.researchgate.net/publication/335691207_Wind_Direction_and_Speed_Prediction_using_Machine_Learning (Accessed: November 17, 2022).

- Second Literature:

  Zareipour, H. *et al.* (2010) *A review of wind power and wind speed forecasting methods with different time horizons*, *Research Gate*. The University of Calgary. Available at: https://www.researchgate.net/publication/224188805_A_review_of_wind_power_and_wind_speed_forecasting_methods_with_different_time_horizons (Accessed: November 17, 2022).

- Third Literature:

  Bhaskar, M., Jain, A. and Naram, S. (2010) *(PDF) wind speed forecasting: Present status - researchgate.net*, *Wind speed forecasting: Present status*. International Institute of Information Technology. Available at: https://www.researchgate.net/publication/224205057_Wind_speed_forecasting_Present_status (Accessed: November 17, 2022).

- 4th Literature:

  Kabošová, L., Chronis, A. and Galanos, T. (2022) *Fast wind prediction incorporated in Urban City Planning*, *Research Gate*. Lenka Kabošová. Available at: https://www.researchgate.net/publication/363547984_Fast_wind_prediction_incorporated_in_urban_city_planning (Accessed: November 19, 2022).

- 5th Literature:

  Upendran, V. *et al.* (2020) *Solar wind prediction using Deep Learning*, *Research Gate*. Vishal Upendran. Available at: https://www.researchgate.net/publication/342100707_Solar_Wind_Prediction_Using_Deep_Learning (Accessed: November 19, 2022).

- 6th Literature:

  Johnson, R. *et al.* (2022) *Physics-informed neural networks for solar wind prediction*, *Research Gate*. Utah State University, Department of Computer Science, Logan, UT, USA, 84322. Available at: https://www.researchgate.net/publication/363250625_Physics-Informed_Neural_Networks_For_Solar_Wind_Prediction (Accessed: November 19, 2022).

- 7th Literature:

  Gillet, Y. and Mori, R. (2021) *ETA and Wind Prediction Accuracy Improvement Using Numerical Weather Forecast and Aircraft Surveillance Data*, *Research Gate*. Yohann Gillet. Available at: https://www.researchgate.net/publication/353532422_ETA_and_Wind_Prediction_Accuracy_Improvement_Using_Numerical_Weather_Forecast_and_Aircraft_Surveillance_Data (Accessed: November 19, 2022).

- 8th Literature:

  Liu, Z. and Ware, T. (2022) *Capturing Spatial Influence in Wind Prediction with a Graph Convolutional Neural Network*, *Research Gate*. Zeyi Liu. Available at: https://www.researchgate.net/publication/359303055_Capturing_Spatial_Influence_in_Wind_Prediction_With_a_Graph_Convolutional_Neural_Network (Accessed: November 19, 2022).

- 9th Literature:

  Manero, J. and Béjar, J. (2021) *Forecastability Measures that Describe the Complexity of a Site for Deep Learning Wind Predictions*, *Research Gate*. Universitat Politècnica de Catalunya. Available at: https://www.researchgate.net/publication/355189377_Forecastability_Measures_that_Describe_the_Complexity_of_a_Site_for_Deep_Learning_Wind_Predictions (Accessed: November 19, 2022).

- 10th Literature:

  Lee, S.W. *et al.* (2013) *Wind prediction in Malaysia*, *Research Gate*. S. W. Lee. Available at: https://www.researchgate.net/publication/286112020_Wind_Prediction_in_Malaysia (Accessed: November 19, 2022).

College of Computing & Engineering

University of Jeddah

كلية علوم وهندسة الحاسب

جامعة جدة

2022

جامعة جدة

University of Jeddah

Chapter #3

3.1 Introduction

We will discuss data collection and where we got the data from in chapter 3, as well as exploratory data analysis, focusing on getting reliable and authentic data based on our expertise.

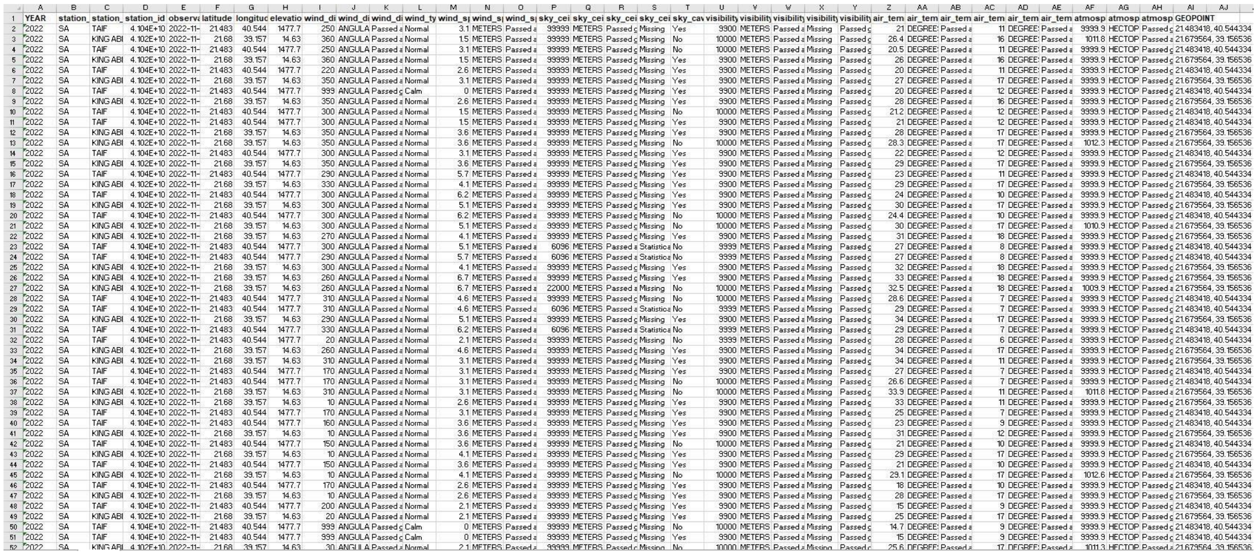
3.2 Data collection

As a first step, we searched the internet for ready-to-use datasets, and we found A data set that meets our needs, the dataset is from King Abdullah Petroleum Studies and Research Center (Saudi Arabia Hourly Climate Integrated Surface Data – KAPSARC Data Portal).

* *Saudi Arabia hourly Climate Integrated Surface Data* (2018) *KAPSARC Data Portal*. National Centers for Environmental Information. Available at: https://datasource.kapsarc.org/explore/dataset/saudi-hourly-weather-data/export/?disjunctive.station_name (Accessed: November 17, 2022).

3.3 Data description

Our project will use the following dataset:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1	YEAR	station_	station_	station_id	observa	latitude	longituc	elevatio	wind_di	wind_di	wind_di	wind_ty	wind_sp	wind_s	wind_s	sky_cei	sky_cei	sky_cei	sky_cei	sky_cav	visibility	visibility	visibility	visibility	visibility	air_tem	air_tem	air_tem	air_tem	air_tem	air_tem	atmosp	atmosp	atmosp	GEOPOINT	
2	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	250	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	21	DEGREE	Passed a	11	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
3	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	360	ANGULA	Passed a	Normal	1.5	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	26.4	DEGREE	Passed a	16	DEGREE	Passed a	1011.8	HECTOP	Passed a	21.679564, 39.156536	
4	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	250	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	20.5	DEGREE	Passed a	11	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
5	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	360	ANGULA	Passed a	Normal	1.5	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	26	DEGREE	Passed a	16	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
6	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	220	ANGULA	Passed a	Normal	2.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	20	DEGREE	Passed a	11	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
7	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	350	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	27	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
8	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	999	ANGULA	Passed g	Calm	0	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	20	DEGREE	Passed a	12	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
9	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	350	ANGULA	Passed a	Normal	2.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	28	DEGREE	Passed a	16	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
10	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	300	ANGULA	Passed a	Normal	1.5	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	21.2	DEGREE	Passed a	12	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
11	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	300	ANGULA	Passed a	Normal	1.5	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	21	DEGREE	Passed a	12	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
12	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	350	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	28	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
13	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	350	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	28.3	DEGREE	Passed a	17	DEGREE	Passed a	1012.3	HECTOP	Passed a	21.679564, 39.156536	
14	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	300	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	22	DEGREE	Passed a	12	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
15	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	350	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	29	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
16	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	290	ANGULA	Passed a	Normal	5.7	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	23	DEGREE	Passed a	11	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
17	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	330	ANGULA	Passed a	Normal	4.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	29	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
18	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	300	ANGULA	Passed a	Normal	6.2	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	24	DEGREE	Passed a	10	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
19	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	300	ANGULA	Passed a	Normal	5.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	30	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
20	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	300	ANGULA	Passed a	Normal	6.2	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	24.4	DEGREE	Passed a	10	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
21	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	300	ANGULA	Passed a	Normal	5.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	30	DEGREE	Passed a	17	DEGREE	Passed a	1010.9	HECTOP	Passed a	21.679564, 39.156536	
22	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	270	ANGULA	Passed a	Normal	4.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	31	DEGREE	Passed a	18	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
23	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	300	ANGULA	Passed a	Normal	5.1	METERS	Passed a	6096	METERS	Passed a	Statistica	No	9999	METERS	Passed a	Missing	Passed g	27	DEGREE	Passed a	8	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
24	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	290	ANGULA	Passed a	Normal	5.7	METERS	Passed a	6096	METERS	Passed a	Statistica	No	9999	METERS	Passed a	Missing	Passed g	27	DEGREE	Passed a	8	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
25	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	300	ANGULA	Passed a	Normal	4.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	32	DEGREE	Passed a	18	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
26	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	260	ANGULA	Passed a	Normal	6.7	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	33	DEGREE	Passed a	18	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
27	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	260	ANGULA	Passed a	Normal	6.7	METERS	Passed a	22000	METERS	Passed a	Missing	No	10000	METERS	Passed a	Missing	Passed g	32.5	DEGREE	Passed a	18	DEGREE	Passed a	1009.9	HECTOP	Passed a	21.679564, 39.156536	
28	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	310	ANGULA	Passed a	Normal	4.6	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	28.6	DEGREE	Passed a	7	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
29	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	310	ANGULA	Passed a	Normal	4.6	METERS	Passed a	6096	METERS	Passed a	Statistica	No	9999	METERS	Passed a	Missing	Passed g	29	DEGREE	Passed a	7	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
30	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	290	ANGULA	Passed a	Normal	5.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	34	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
31	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	330	ANGULA	Passed a	Normal	6.2	METERS	Passed a	6096	METERS	Passed a	Statistica	No	9999	METERS	Passed a	Missing	Passed g	29	DEGREE	Passed a	7	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
32	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	20	ANGULA	Passed a	Normal	2.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	9999	METERS	Passed a	Missing	Passed g	28	DEGREE	Passed a	6	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
33	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	260	ANGULA	Passed a	Normal	4.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	34	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
34	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	310	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	34	DEGREE	Passed a	11	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
35	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	170	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	27	DEGREE	Passed a	7	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
36	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	170	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	26.6	DEGREE	Passed a	7	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
37	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	310	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	33.9	DEGREE	Passed a	11	DEGREE	Passed a	1011.8	HECTOP	Passed a	21.679564, 39.156536	
38	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	10	ANGULA	Passed a	Normal	2.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	33	DEGREE	Passed a	11	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
39	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	170	ANGULA	Passed a	Normal	3.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	25	DEGREE	Passed a	7	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
40	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	160	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	23	DEGREE	Passed a	9	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
41	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	10	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	31	DEGREE	Passed a	12	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
42	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	150	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	21	DEGREE	Passed a	10	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
43	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	10	ANGULA	Passed a	Normal	4.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	29	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
44	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	150	ANGULA	Passed a	Normal	3.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	21	DEGREE	Passed a	10	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
45	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	10	ANGULA	Passed a	Normal	4.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	29.1	DEGREE	Passed a	17	DEGREE	Passed a	1012.6	HECTOP	Passed a	21.679564, 39.156536	
46	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	170	ANGULA	Passed a	Normal	2.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	18	DEGREE	Passed a	10	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
47	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	10	ANGULA	Passed a	Normal	2.6	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	28	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
48	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	200	ANGULA	Passed a	Normal	2.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	15	DEGREE	Passed a	9	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
49	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	20	ANGULA	Passed a	Normal	2.1	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	25	DEGREE	Passed a	17	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.679564, 39.156536	
50	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	999	ANGULA	Passed g	Calm	0	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	14.7	DEGREE	Passed a	9	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
51	2022	SA	TAIF	4.104E+10	2022-11-	21.483	40.544	1477.7	999	ANGULA	Passed g	Calm	0	METERS	Passed a	99999	METERS	Passed g	Missing	Yes	9900	METERS	Passed a	Missing	Passed g	15	DEGREE	Passed a	9	DEGREE	Passed a	9999.9	HECTOP	Passed g	21.483418, 40.544334	
52	2022	SA	KING ABI	4.102E+10	2022-11-	21.68	39.157	14.63	30	ANGULA	Passed a	Normal	2.1	METERS	Passed a	99999	METERS	Passed g	Missing	No	10000	METERS	Passed a	Missing	Passed g	25.6	DEGREE	Passed a	17	DEGREE	Passed a	1011.3	HECTOP	Passed a	21.679564, 39.156536	

Figure 1 Dataset

Our dataset includes **(35) columns** and **(122816) rows** and contains 3 stations in the west of the Kingdom of Saudi Arabia **(Jeddah, Arafat, and Taif)** the dataset collection started from **2018/1/1 to 2022/11/11**. Which means before 5 years until the end of this year 2022.

The column in our dataset contains the observation date, observation year, station name, latitude, longitude, elevation, wind direction angle, wind type, wind speed rate, wind speed rate unit, sky ceiling height, sky ceiling height units, sky cavok, visibility distance, visibility distance units, air temperature, air temperature units, air temperature dew point, air temperature dew point units, atmospheric sea level pressure, atmospheric sea level pressure units, and Geopoint.

3.4 Exploratory data analytics

We may remove some columns that we see is not helpful from the dataset that we found, we want to make sure that all the columns that we have are useful and related to our project.

For example, we will remove longitude and latitude because we are not locking in locations, that’s the reason for removing these columns.

We will keep looking if there are any useless columns in the dataset.

3.5 Conclusion

To summarize, this project's pre-processing phase is a challenging yet fun process because it is an integral part of the project. Any errors during this phase will affect our results, so we need to handle the data in a way that maintains its integrity during this phase.

College of Computing & Engineering

University of Jeddah

كلية علوم وهندسة الحاسب

جامعة جدة

2022

Chapter #4

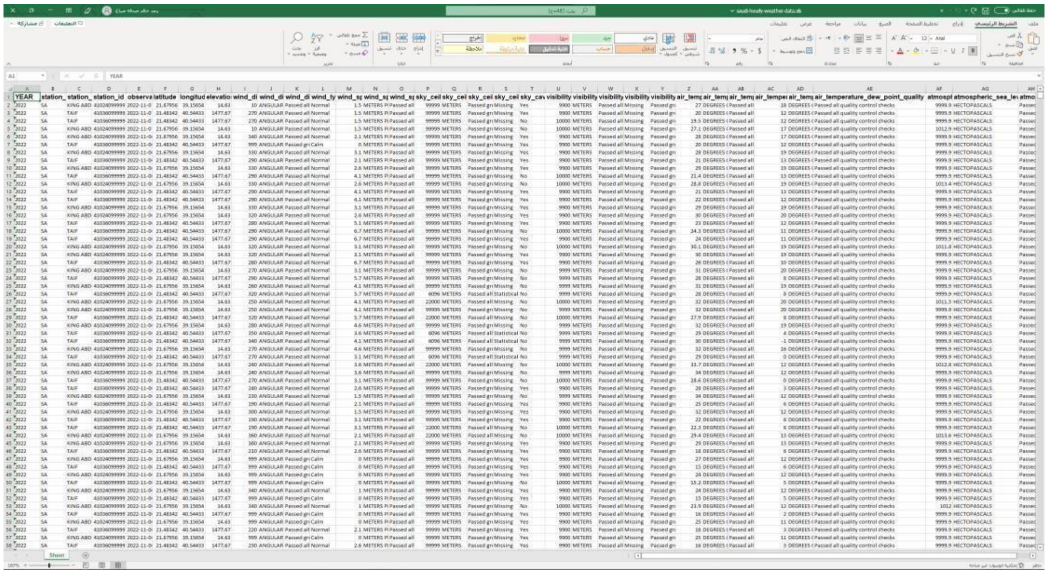
4.1 Introduction

We will preprocess the dataset we have to collect, and what are the problems we faced,and how we handle it by using the techniques that we discussed in the preview chapters.

4.2 Problems with Dataset

We have found many problems in our dataset, the problems we have indicated were

- Useless columns
- Wrong values
- Wrong units
- Missing columns
- Wrong types



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	YEAR	station_	station_	station_id	observa	latitude	longitud	elevatio	wind_di	wind_di	wind_di	wind_ty	wind_sp	wind_sp	wind_sp	sky_ceil	sky_ceil	sky_ceil	sky_ceil	sky_cav	visibility	visibility	visibility	visibility	visibility	air_tem	air_tem	air_tem	air_temper	air_tem	air_temperature_dew_point_quality	atmospl	atmospheric_sea_le	atmo
2	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	10	ANGULAR	Passed all	Normal	1.5	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	27	DEGREES (	Passed all	18	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
3	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	270	ANGULAR	Passed all	Normal	1.5	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	20	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
4	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	270	ANGULAR	Passed all	Normal	1.5	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	19.5	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
5	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	10	ANGULAR	Passed all	Normal	1.5	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	27.1	DEGREES (	Passed all	17	DEGREES (	Passed all quality control checks	1012.9	HECTOPASCALS	Passec
6	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	340	ANGULAR	Passed all	Normal	2.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	28	DEGREES (	Passed all	17	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
7	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	20	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
8	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	330	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	28	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
9	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	2.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	21	DEGREES (	Passed all	13	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
10	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	330	ANGULAR	Passed all	Normal	2.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	29	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
11	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	21.4	DEGREES (	Passed all	13	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
12	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	330	ANGULAR	Passed all	Normal	2.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	28.8	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	1013.4	HECTOPASCALS	Passec
13	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	21	DEGREES (	Passed all	13	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
14	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	22	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
15	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	330	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	29	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
16	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	320	ANGULAR	Passed all	Normal	2.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	30	DEGREES (	Passed all	20	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
17	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	280	ANGULAR	Passed all	Normal	5.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	23	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
18	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	6.7	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	24.3	DEGREES (	Passed all	11	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
19	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	6.7	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	24	DEGREES (	Passed all	11	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
20	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	320	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	30.1	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	1011.8	HECTOPASCALS	Passec
21	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	320	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	30	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
22	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	280	ANGULAR	Passed all	Normal	6.7	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	26	DEGREES (	Passed all	10	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
23	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	270	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	31	DEGREES (	Passed all	20	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
24	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	290	ANGULAR	Passed all	Normal	5.7	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	28	DEGREES (	Passed all	8	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
25	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	260	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	31	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
26	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	320	ANGULAR	Passed all	Normal	5.7	METERS PI	Passed all	6096	METERS	Passed all	Statistical	No	9999	METERS	Passed all	Missing	Passed gn	28	DEGREES (	Passed all	8	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
27	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	250	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	22000	METERS	Passed all	Missing	No	10000	METERS	Passed all	Missing	Passed gn	32	DEGREES (	Passed all	20	DEGREES (	Passed all quality control checks	1011.5	HECTOPASCALS	Passec
28	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	250	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	32	DEGREES (	Passed all	20	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
29	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	320	ANGULAR	Passed all	Normal	5.7	METERS PI	Passed all	22000	METERS	Passed all	Missing	No	10000	METERS	Passed all	Missing	Passed gn	27.9	DEGREES (	Passed all	8	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
30	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	280	ANGULAR	Passed all	Normal	4.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	32	DEGREES (	Passed all	19	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
31	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	350	ANGULAR	Passed all	Normal	3.6	METERS PI	Passed all	6096	METERS	Passed all	Statistical	No	9999	METERS	Passed all	Missing	Passed gn	29	DEGREES (	Passed all	4	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
32	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	340	ANGULAR	Passed all	Normal	4.1	METERS PI	Passed all	6096	METERS	Passed all	Statistical	No	9999	METERS	Passed all	Missing	Passed gn	30	DEGREES (	Passed all	-1	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
33	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	270	ANGULAR	Passed all	Normal	4.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	32	DEGREES (	Passed all	16	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
34	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	270	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	6096	METERS	Passed all	Statistical	No	9999	METERS	Passed all	Missing	Passed gn	29	DEGREES (	Passed all	0	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
35	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	240	ANGULAR	Passed all	Normal	3.6	METERS PI	Passed all	22000	METERS	Passed all	Missing	No	10000	METERS	Passed all	Missing	Passed gn	33.7	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	1012.8	HECTOPASCALS	Passec
36	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	240	ANGULAR	Passed all	Normal	3.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	34	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
37	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	270	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	28.6	DEGREES (	Passed all	0	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
38	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	240	ANGULAR	Passed all	Normal	2.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	28	DEGREES (	Passed all	3	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
39	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	230	ANGULAR	Passed all	Normal	1.5	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	9999	METERS	Passed all	Missing	Passed gn	34	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
40	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	190	ANGULAR	Passed all	Normal	2.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	25	DEGREES (	Passed all	6	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
41	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	300	ANGULAR	Passed all	Normal	1.5	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	32	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
42	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	190	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	22	DEGREES (	Passed all	8	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
43	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	190	ANGULAR	Passed all	Normal	3.1	METERS PI	Passed all	22000	METERS	Passed all	Missing	No	10000	METERS	Passed all	Missing	Passed gn	22.3	DEGREES (	Passed all	8	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
44	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	360	ANGULAR	Passed all	Normal	2.1	METERS PI	Passed all	22000	METERS	Passed all	Missing	No	10000	METERS	Passed all	Missing	Passed gn	29.4	DEGREES (	Passed all	13	DEGREES (	Passed all quality control checks	1013.6	HECTOPASCALS	Passec
45	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	360	ANGULAR	Passed all	Normal	2.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	29	DEGREES (	Passed all	13	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
46	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	210	ANGULAR	Passed all	Normal	2.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	18	DEGREES (	Passed all	6	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
47	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	27	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
48	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	15	DEGREES (	Passed all	6	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
49	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	24	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
50	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	15.2	DEGREES (	Passed all	5	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
51	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	340	ANGULAR	Passed all	Normal	1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	24	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
52	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	15	DEGREES (	Passed all	5	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
53	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	340	ANGULAR	Passed all	Normal	1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	No	10000	METERS	Passed all	Missing	Passed gn	23.9	DEGREES (	Passed all	12	DEGREES (	Passed all quality control checks	1012	HECTOPASCALS	Passec
54	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	16	DEGREES (	Passed all	2	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
55	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	25	DEGREES (	Passed all	11	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
56	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	220	ANGULAR	Passed all	Normal	2.1	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	16	DEGREES (	Passed all	3	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
57	2022	SA	KING ABD	41024099999	2022-11-0	21.67956	39.15654	14.63	999	ANGULAR	Passed gr	Calm	0	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	25	DEGREES (	Passed all	11	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec
58	2022	SA	TAIF	41036099999	2022-11-0	21.48342	40.54433	1477.67	230	ANGULAR	Passed all	Normal	2.6	METERS PI	Passed all	99999	METERS	Passed gn	Missing	Yes	9900	METERS	Passed all	Missing	Passed gn	16	DEGREES (	Passed all	3	DEGREES (	Passed all quality control checks	9999.9	HECTOPASCALS	Passec

Figure 2 Dataset

First of all, the problem we faced started with a useless column, and wrong values in manyrows, we also found incorrect units in multiple columns, and the result for wind_type was horrific and completely inaccurate

4.3 Data Preparation and Preprocessing

We will be removing the unnecessary columns which are:

Station_country
Latitude
Longitude
Sky_ceiling_height
Wind_direction_angle_units
Wind_direction_quality
Wind_speed_rate_units
Wind_speed_quality
Sky_ceiling_height_units
Sky_ceiling_quality
Sky_ceiling_determination
Visibility_distance_quality
Visibility_variablity
Visibility_variability_quality
Air_temperature_units
Air_temperature_quality
Air_temperature_daw_point_units
Air_temperature_dew_point_quality
Atmospheric_sea_level_pressure
Atmospheric_sea_level_pressure_units
Atmospheric_sea_level_pressure_quality
After removing the unnecessary columns, the dataset includes 13 columns instead 35 columns and 122943 rows.

there were many wrong values in wind speed rate, visibility distance, air temperature, andwind direction angle, so we dropped the wrong values continues (999.9, 999999, 999)

This row was dropped because there were so many incorrect values, except for taking the average.

The average will not be useful for us because the wrong values in each column were almost 30.000 rows.

Now our dataset after dropping the wrong values continues 95115 rows and 13 columns

We also changed the unit of the wind speed rate from m/s to km/h, to make our forecasting more accurate since we are predicting a Saudi Arabia weather dataset, as well we changed thevisibility distance rate from m to km.

the main purpose of changing these units is because it’s certified by the national center of meteorology in the Kingdom of Saudi Arabia.

We also found the visibility distance close to 10 and the accurate number certified its integer, so every row containing 9.999 was converted to 10 to make the prediction moreaccurate.

We changed the wind types based on the wind speed in our dataset to the shown standard global rate according to Wikipedia.

If the wind speed was less than 1, the condition is ‘**Calm**’
If it’s less than 5 the condition is ‘**Light air**’
If it’s less than 12 the condition is ‘**Light Breeze**’
If it’s less than 20 the condition is ‘**Gentle Breeze**’
If it’s less than 29 the condition is ‘**Moderate Breeze**’
If it’s less than 39 the condition is ‘**Fresh Breeze**’
If it’s less than 50 the condition is ‘**Strong Breeze**’
If it’s less than 62 the condition is ‘**Near Gale**’
If it’s less than 75 the condition is ‘**Gale**’
If it’s less than 89 the condition is ‘**Strong Gale’**
If it’s less than 103 the condition is ‘**Storm**’
If it’s less than 118 the condition is ‘**Violent Storm**’
More than 118 conditions will be ‘**Hurricane**’

We test our new wind type conditions best on the wind speed and the max condition on the 3 years was ‘Strong gale’.

Our wind-type locks are in good shape now after the preprocessing and will be more helpful in ourforecasting.

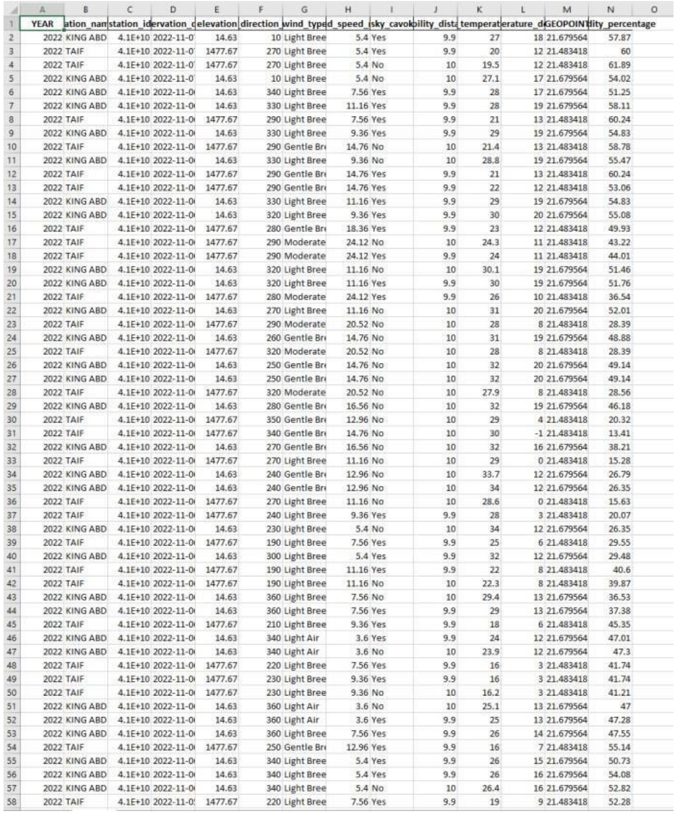
The last thing we have done is add a new column 'Humidity' by calculating it based on the air temperature and the Dew point by using constants we fowled. We think we might use it to see if it affects the wind speed or not

$$D_p = \frac{\lambda \times \left(\ln\left(\frac{RH}{100}\right) + \frac{\beta\, T}{\lambda + T} \right)}{\beta - \left(\ln\left(\frac{RH}{100}\right) + \frac{\beta\, T}{\lambda + T} \right)}$$

*Singh, P. (2022) *Relative humidity calculator*, *Omni Calculator*. Omni Calculator. Available at: https://www.omnicalculator.com/physics/relative-humidity (Accessed: November 17, 2022).

4.4 Conclusion

In this chapter, we handled our dataset by using python techniques and libraries, which made our preprocessing easier. from dealing with wrong values and adding important columns and more.
after we finished cleaning our dataset, we export our file from CVS to XLSX to be more suitable.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	YEAR	ation_nam	station_id	ervation_d	elevation	direction_	wind_type	d_speed_	sky_cavok	bility_dist	temperat	erature_d	GEOPOINT	dity_percentage	
2	2022	KING ABD	4.1E+10	2022-11-0	14.63	10	Light Bree	5.4	Yes	9.9	27	18	21.679564	57.87	
3	2022	TAIF	4.1E+10	2022-11-0	1477.67	270	Light Bree	5.4	Yes	9.9	20	12	21.483418	60	
4	2022	TAIF	4.1E+10	2022-11-0	1477.67	270	Light Bree	5.4	No	10	19.5	12	21.483418	61.89	
5	2022	KING ABD	4.1E+10	2022-11-0	14.63	10	Light Bree	5.4	No	10	27.1	17	21.679564	54.02	
6	2022	KING ABD	4.1E+10	2022-11-0	14.63	340	Light Bree	7.56	Yes	9.9	28	17	21.679564	51.25	
7	2022	KING ABD	4.1E+10	2022-11-0	14.63	330	Light Bree	11.16	Yes	9.9	28	19	21.679564	58.11	
8	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Light Bree	7.56	Yes	9.9	21	13	21.483418	60.24	
9	2022	KING ABD	4.1E+10	2022-11-0	14.63	330	Light Bree	9.36	Yes	9.9	29	19	21.679564	54.83	
10	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Gentle Br	14.76	No	10	21.4	13	21.483418	58.78	
11	2022	KING ABD	4.1E+10	2022-11-0	14.63	330	Light Bree	9.36	No	10	28.8	19	21.679564	55.47	
12	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Gentle Br	14.76	Yes	9.9	21	13	21.483418	60.24	
13	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Gentle Br	14.76	Yes	9.9	22	12	21.483418	53.06	
14	2022	KING ABD	4.1E+10	2022-11-0	14.63	330	Light Bree	11.16	Yes	9.9	29	19	21.679564	54.83	
15	2022	KING ABD	4.1E+10	2022-11-0	14.63	320	Light Bree	9.36	Yes	9.9	30	20	21.679564	55.08	
16	2022	TAIF	4.1E+10	2022-11-0	1477.67	280	Gentle Br	18.36	Yes	9.9	23	12	21.483418	49.93	
17	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Moderate	24.12	No	10	24.3	11	21.483418	43.22	
18	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Moderate	24.12	Yes	9.9	24	11	21.483418	44.01	
19	2022	KING ABD	4.1E+10	2022-11-0	14.63	320	Light Bree	11.16	No	10	30.1	19	21.679564	51.46	
20	2022	KING ABD	4.1E+10	2022-11-0	14.63	320	Light Bree	11.16	Yes	9.9	30	19	21.679564	51.76	
21	2022	TAIF	4.1E+10	2022-11-0	1477.67	280	Moderate	24.12	Yes	9.9	26	10	21.483418	36.54	
22	2022	KING ABD	4.1E+10	2022-11-0	14.63	270	Light Bree	11.16	No	10	31	20	21.679564	52.01	
23	2022	TAIF	4.1E+10	2022-11-0	1477.67	290	Moderate	20.52	No	10	28	8	21.483418	28.39	
24	2022	KING ABD	4.1E+10	2022-11-0	14.63	260	Gentle Br	14.76	No	10	31	19	21.679564	48.88	
25	2022	TAIF	4.1E+10	2022-11-0	1477.67	320	Moderate	20.52	No	10	28	8	21.483418	28.39	
26	2022	KING ABD	4.1E+10	2022-11-0	14.63	250	Gentle Br	14.76	No	10	32	20	21.679564	49.14	
27	2022	KING ABD	4.1E+10	2022-11-0	14.63	250	Gentle Br	14.76	No	10	32	20	21.679564	49.14	
28	2022	TAIF	4.1E+10	2022-11-0	1477.67	320	Moderate	20.52	No	10	27.9	8	21.483418	28.56	
29	2022	KING ABD	4.1E+10	2022-11-0	14.63	280	Gentle Br	16.56	No	10	32	19	21.679564	46.18	
30	2022	TAIF	4.1E+10	2022-11-0	1477.67	350	Gentle Br	12.96	No	10	29	4	21.483418	20.32	
31	2022	TAIF	4.1E+10	2022-11-0	1477.67	340	Gentle Br	14.76	No	10	30	-1	21.483418	13.41	
32	2022	KING ABD	4.1E+10	2022-11-0	14.63	270	Gentle Br	16.56	No	10	32	16	21.679564	38.21	
33	2022	TAIF	4.1E+10	2022-11-0	1477.67	270	Light Bree	11.16	No	10	29	0	21.483418	15.28	
34	2022	KING ABD	4.1E+10	2022-11-0	14.63	240	Gentle Br	12.96	No	10	33.7	12	21.679564	26.79	
35	2022	KING ABD	4.1E+10	2022-11-0	14.63	240	Gentle Br	12.96	No	10	34	12	21.679564	26.35	
36	2022	TAIF	4.1E+10	2022-11-0	1477.67	270	Light Bree	11.16	No	10	28.6	0	21.483418	15.63	
37	2022	TAIF	4.1E+10	2022-11-0	1477.67	240	Light Bree	9.36	Yes	9.9	28	3	21.483418	20.07	
38	2022	KING ABD	4.1E+10	2022-11-0	14.63	230	Light Bree	5.4	No	10	34	12	21.679564	26.35	
39	2022	TAIF	4.1E+10	2022-11-0	1477.67	190	Light Bree	7.56	Yes	9.9	25	6	21.483418	29.55	
40	2022	KING ABD	4.1E+10	2022-11-0	14.63	300	Light Bree	5.4	Yes	9.9	32	12	21.679564	29.48	
41	2022	TAIF	4.1E+10	2022-11-0	1477.67	190	Light Bree	11.16	Yes	9.9	22	8	21.483418	40.6	
42	2022	TAIF	4.1E+10	2022-11-0	1477.67	190	Light Bree	11.16	No	10	22.3	8	21.483418	39.87	
43	2022	KING ABD	4.1E+10	2022-11-0	14.63	360	Light Bree	7.56	No	10	29.4	13	21.679564	36.53	
44	2022	KING ABD	4.1E+10	2022-11-0	14.63	360	Light Bree	7.56	Yes	9.9	29	13	21.679564	37.38	
45	2022	TAIF	4.1E+10	2022-11-0	1477.67	210	Light Bree	9.36	Yes	9.9	18	6	21.483418	45.35	
46	2022	KING ABD	4.1E+10	2022-11-0	14.63	340	Light Air	3.6	Yes	9.9	24	12	21.679564	47.01	
47	2022	KING ABD	4.1E+10	2022-11-0	14.63	340	Light Air	3.6	No	10	23.9	12	21.679564	47.3	
48	2022	TAIF	4.1E+10	2022-11-0	1477.67	220	Light Bree	7.56	Yes	9.9	16	3	21.483418	41.74	
49	2022	TAIF	4.1E+10	2022-11-0	1477.67	230	Light Bree	9.36	Yes	9.9	16	3	21.483418	41.74	
50	2022	TAIF	4.1E+10	2022-11-0	1477.67	230	Light Bree	9.36	No	10	16.2	3	21.483418	41.21	
51	2022	KING ABD	4.1E+10	2022-11-0	14.63	360	Light Air	3.6	No	10	25.1	13	21.679564	47	
52	2022	KING ABD	4.1E+10	2022-11-0	14.63	360	Light Air	3.6	Yes	9.9	25	13	21.679564	47.28	
53	2022	KING ABD	4.1E+10	2022-11-0	14.63	360	Light Bree	7.56	Yes	9.9	26	14	21.679564	47.55	
54	2022	TAIF	4.1E+10	2022-11-0	1477.67	250	Gentle Br	12.96	Yes	9.9	16	7	21.483418	55.14	
55	2022	KING ABD	4.1E+10	2022-11-0	14.63	340	Light Bree	5.4	Yes	9.9	26	15	21.679564	50.73	
56	2022	KING ABD	4.1E+10	2022-11-0	14.63	340	Light Bree	5.4	Yes	9.9	26	16	21.679564	54.08	
57	2022	KING ABD	4.1E+10	2022-11-0	14.63	340	Light Bree	5.4	No	10	26.4	16	21.679564	52.82	
58	2022	TAIF	4.1E+10	2022-11-0	1477.67	220	Light Bree	7.56	Yes	9.9	19	9	21.483418	52.28	

After preprocessing our dataset has reached 95115 rows and 14 columns instead of 122943 rows and 35 columns.

References (*):

- R, S.E. (2022) *Random Forest: Introduction to random forest algorithm*, *Analytics Vidhya*. Sruthi E R. Available at: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/ (Accessed: November 17, 2022).

- Stecanella, B. (2017) *Support Vector Machines (SVM) algorithm explained*, *MonkeyLearn Blog*. Bruno Stecanella. Available at: https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/ (Accessed: November 17, 2022).

- Lawton, G., Burns, E. and Rosencrance, L. (2022) *What is logistic regression? - definition from Searchbusinessanalytics*, *SearchBusinessAnalytics*. TechTarget. Available at: https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression (Accessed: November 17, 2022).

- Resources, M. (2016) *Pandas: Python library - mode*, *Mode Resources*. Mode Resources. Available at: https://mode.com/python-tutorial/libraries/pandas/ (Accessed: November 17, 2022).

- w3schools (no date) *Numpy introduction*, *Introduction to NumPy*. w3schools. Available at: https://www.w3schools.com/python/numpy/numpy_intro.asp (Accessed: November 17, 2022).

- First Literature:

Music, E., Halilovic, Jusufovic, Kevric (2018) *Wind direction and speed prediction using machine learning*, *Research Gate*. International BURCH University. Available at: https://www.researchgate.net/publication/335691207_Wind_Direction_and_Speed_Prediction_using_Machine_Learning (Accessed: November 17, 2022).

- Second Literature:

Zareipour, H. *et al.* (2010) *A review of wind power and wind speed forecasting methods with different time horizons*, *Research Gate*. The University of Calgary. Available at: https://www.researchgate.net/publication/224188805_A_review_of_wind_power_and_wind_speed_forecasting_methods_with_different_time_horizons (Accessed: November 17, 2022).

- Third Literature:

Bhaskar, M., Jain, A. and Naram, S. (2010) *(PDF) wind speed forecasting: Present status - researchgate.net*, *Wind speed forecasting: Present status*. International Institute of Information Technology. Available at: https://www.researchgate.net/publication/224205057_Wind_speed_forecasting_Present_status (Accessed: November 17, 2022).

- 4th Literature:

Kabošová, L., Chronis, A. and Galanos, T. (2022) *Fast wind prediction incorporated in Urban City Planning*, *Research Gate*. Lenka Kabošová. Available at: https://www.researchgate.net/publication/363547984_Fast_wind_prediction_incorporated_in_urban_city_planning (Accessed: November 19, 2022).

- 5th Literature:

Upendran, V. *et al.* (2020) *Solar wind prediction using Deep Learning*, *Research Gate*. Vishal Upendran. Available at: https://www.researchgate.net/publication/342100707_Solar_Wind_Prediction_Using_Deep_Learning (Accessed: November 19, 2022).

- 6th Literature:

  Johnson, R. *et al.* (2022) *Physics-informed neural networks for solar wind prediction*, *Research Gate*. Utah State University, Department of Computer Science, Logan, UT, USA, 84322. Available at: https://www.researchgate.net/publication/363250625_Physics-Informed_Neural_Networks_For_Solar_Wind_Prediction (Accessed: November 19, 2022).

- 7th Literature:

  Gillet, Y. and Mori, R. (2021) *ETA and Wind Prediction Accuracy Improvement Using Numerical Weather Forecast and Aircraft Surveillance Data*, *Research Gate*. Yohann Gillet. Available at: https://www.researchgate.net/publication/353532422_ETA_and_Wind_Prediction_Accuracy_Improvement_Using_Numerical_Weather_Forecast_and_Aircraft_Surveillance_Data (Accessed: November 19, 2022).

- 8th Literature:

  Liu, Z. and Ware, T. (2022) *Capturing Spatial Influence in Wind Prediction with a Graph Convolutional Neural Network*, *Research Gate*. Zeyi Liu. Available at: https://www.researchgate.net/publication/359303055_Capturing_Spatial_Influence_in_Wind_Prediction_With_a_Graph_Convolutional_Neural_Network (Accessed: November 19, 2022).

- 9th Literature:

  Manero, J. and Béjar, J. (2021) *Forecastability Measures that Describe the Complexity of a Site for Deep Learning Wind Predictions*, *Research Gate*. Universitat Politècnica de Catalunya. Available at: https://www.researchgate.net/publication/355189377_Forecastability_Measures_that_Describe_the_Complexity_of_a_Site_for_Deep_Learning_Wind_Predictions (Accessed: November 19, 2022).

- 10th Literature:

  Lee, S.W. *et al.* (2013) *Wind prediction in Malaysia*, *Research Gate*. S. W. Lee. Available at: https://www.researchgate.net/publication/286112020_Wind_Prediction_in_Malaysia (Accessed: November 19, 2022).

- *Saudi Arabia hourly Climate Integrated Surface Data* (2018) *KAPSARC Data Portal*. National Centers for Environmental Information. Available at: https://datasource.kapsarc.org/explore/dataset/saudi-hourly-weather-data/export/?disjunctive.station_name (Accessed: November 17, 2022).

- *Singh, P. (2022) *Relative humidity calculator*, *Omni Calculator*. Omni Calculator. Available at: https://www.omnicalculator.com/physics/relative-humidity (Accessed: November 17, 2022

College of Computing & Engineering

University of Jeddah

كلية علوم وهندسة الحاسب

جامعة جدة

2023

Final Senior Project #2 Report

DONE BY:
AHMED ALMDAIFER
ABDULAZIZ MGARRY
RAAD SABBAGH

SUPERVISOR:
DR. OMAR AL-GHUSHAIRY

TABLE OF CONTENT

Chapter 5: Model Building

5.1 Introduction

This chapter discusses the model-building phase of the project. Our discussion will cover the various tools we used and the libraries that we used, as well as how we used the previously mentioned datasets and how we manipulated them to meet our project's goals.

5.2 Experiments Setup and Tools

All our work was done using machine learning algorithms and python libraries such as pandas and NumPy and SciPy. We also used data visualization libraries such as matplotlib, and Seaborn.

Seeing that our data has many features, we will be focusing on wind speed rate and wind speed direction, to give us the answer to our questions (whether the wind speed affects visibility or not and if it affects the air temperature or not)
We also will show if there is any relation between the other features.

5.3 Initial Parameters and Selection Criteria

We will use multiple machine learning algorithms including Decision Tree Regression, Random Forest Regression, XGBoost Regression. The five models will be trained based on wind speed rate, then our machine learning algorithms will be compared to find the most accurate model for our project.

5.4 Conclusion

Our model is very useful for achieving our end goal in this project, which we hope will come in handy in the future for predicting the city of Jeddah, Makkah, and Taif. We can apply this model to many other cities by changing the data, and if the data given to us is accurate and detailed, we can use the same model, and the results will be direct and concise.

Chapter 6: Results and Discussions

6.1 Introduction

In this chapter, we going to apply all the tools and algorithms that we selected to find out which model has the highest accuracy and if the wind speed and direction affect the visibility and temperature or not, and see if it meets our expectations or if it would be different results.

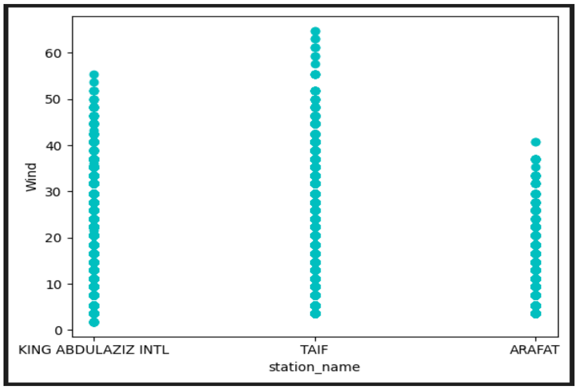
6.2 Performance Evaluation Metrics

We will check the accuracies of the models to find out which one has the highest and lowest accuracies. The performance evaluation metrics will be Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE). We also have R2 Score (pronounced R-Squared Score) for measuring the accuracy directly.

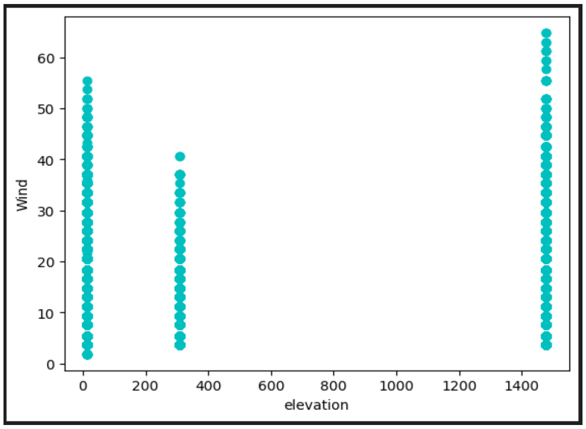
6.3 Experiments Results

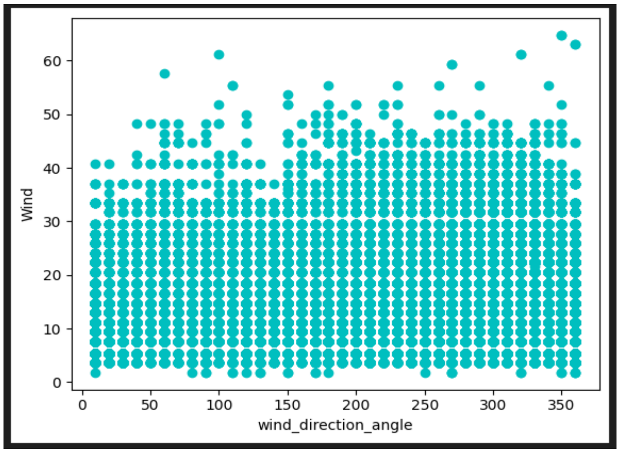
We compared all our features with the wind speed, first of all, we have three stations in our dataset. We split these stations to see which of all stations has the highest wind speed, after that completed Our comparison with all features. Each graph has an explanation for the relation with the wind speed.

- Wind Speed Rate vs station name - as we see in this graph Taif recorded the highest wind speed over the three stations.

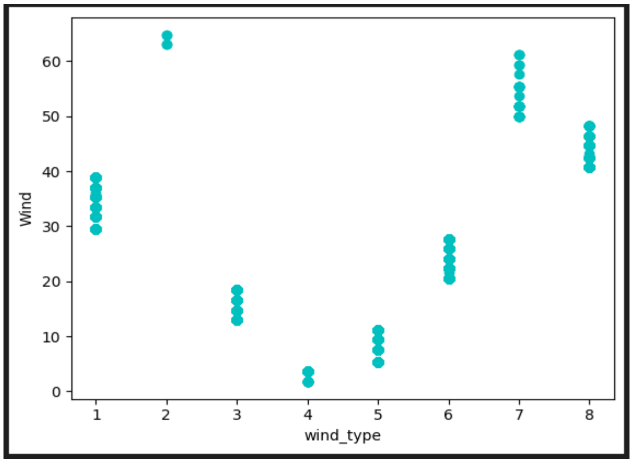


- Wind Speed Rate vs Elevation - When the elevation is very high and very low, the wind speed shows a significant increase. For the range between 400-1200, there is not any wind recorded.

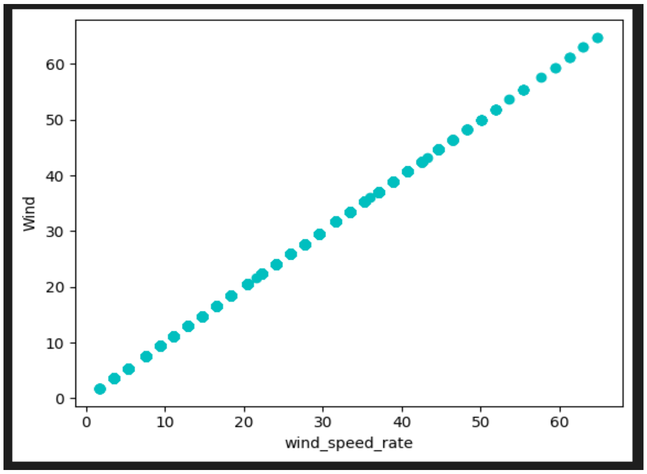




- Wind vs Wind Direction - As you can see, the wind is blowing in all directions and it's not specific to a particular angle.

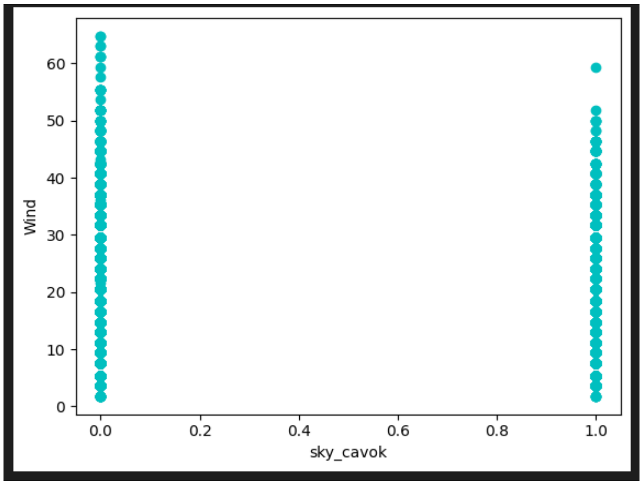


- Wind Speed Rate vs Wind type - 4 type is the slowest wind and the 2 types is the highest wind type.

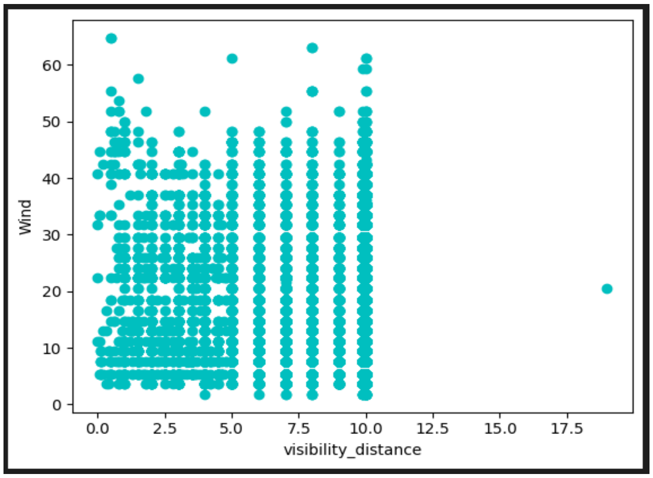


- Wind Speed Rate vs Wind speed rate - It is showing a linear relationship between the two. So, we can say it's highly co-related.

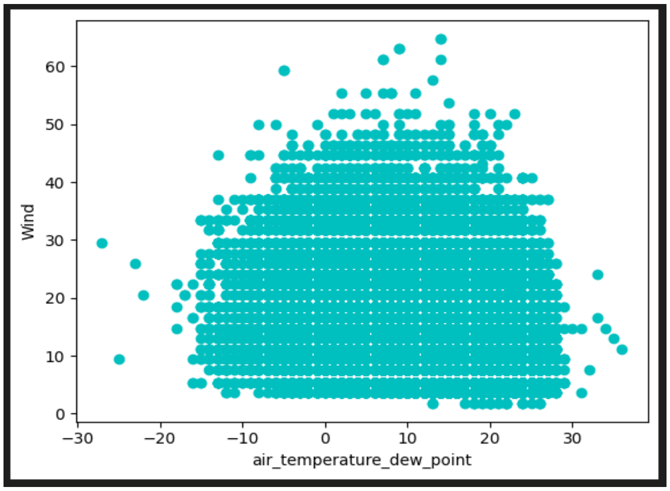
- Wind Speed Rate vs Sky_Cavok, we can't express any idea about this graph.

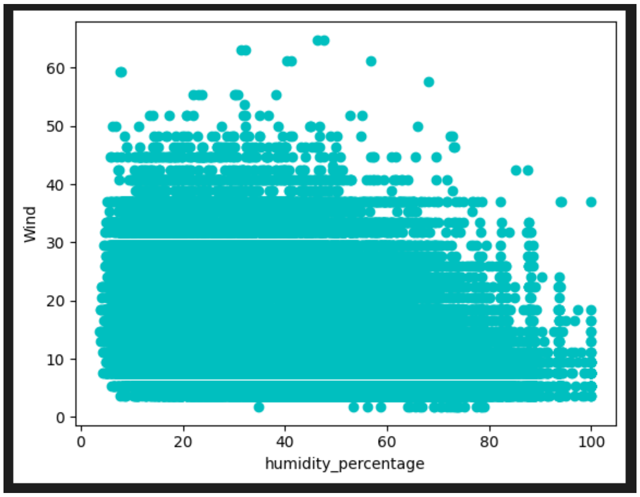


- Wind Speed Rate vs Visibility, When the visibility increases, there are not any records of wind speed. So, we can say that the Wind is limited to specific visibility areas such as 0 to 10. There is an outlier in the graph as well.

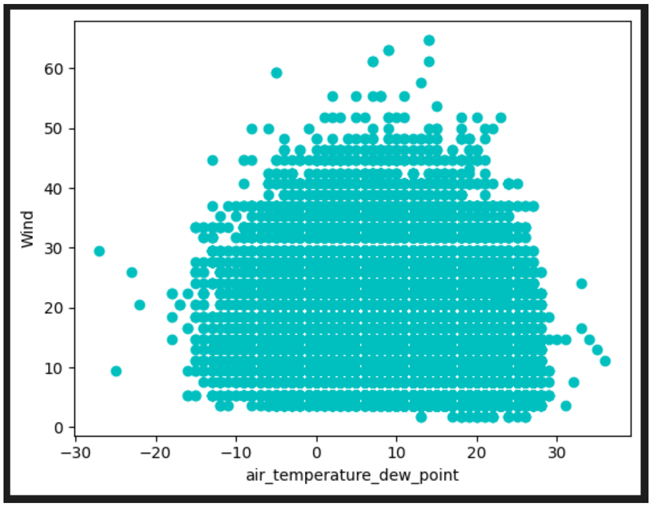


- Wind Speed Rate vs Air_Temperature, we are observing a high wind speed between normal temperatures (20-30) Other temperature levels are not showing a considerable speed.

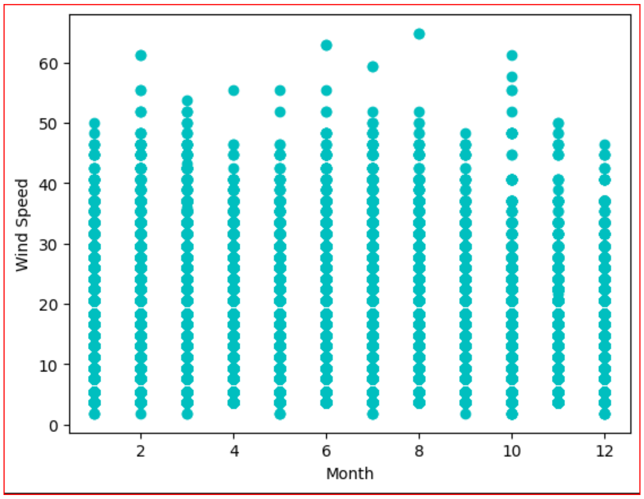




- Wind Speed Rate vs Humidity - When humidity increases, wind speed is getting slow. As the humidity increased up to 100, we can observe a great slowness of the wind speed.



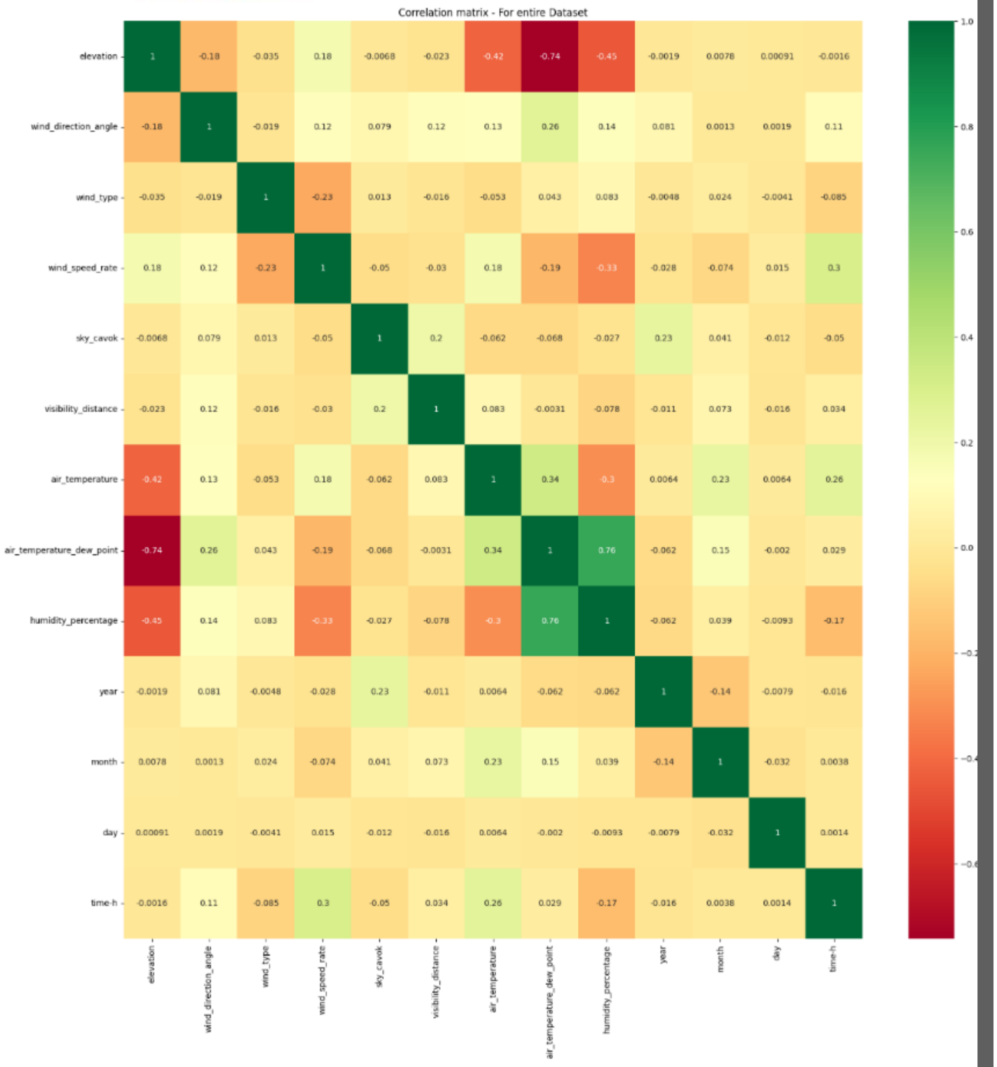
- Wind Speed Rate vs Dew point, this is also similar to the air temperature graph. In between -20 to 30, considerable windscreens can be observed.



- Wind Speed Rate vs Monthe, as we can see there is a relation between wind speed and four months. For the five years most, high winds show up in specific months (2,6,8,10)

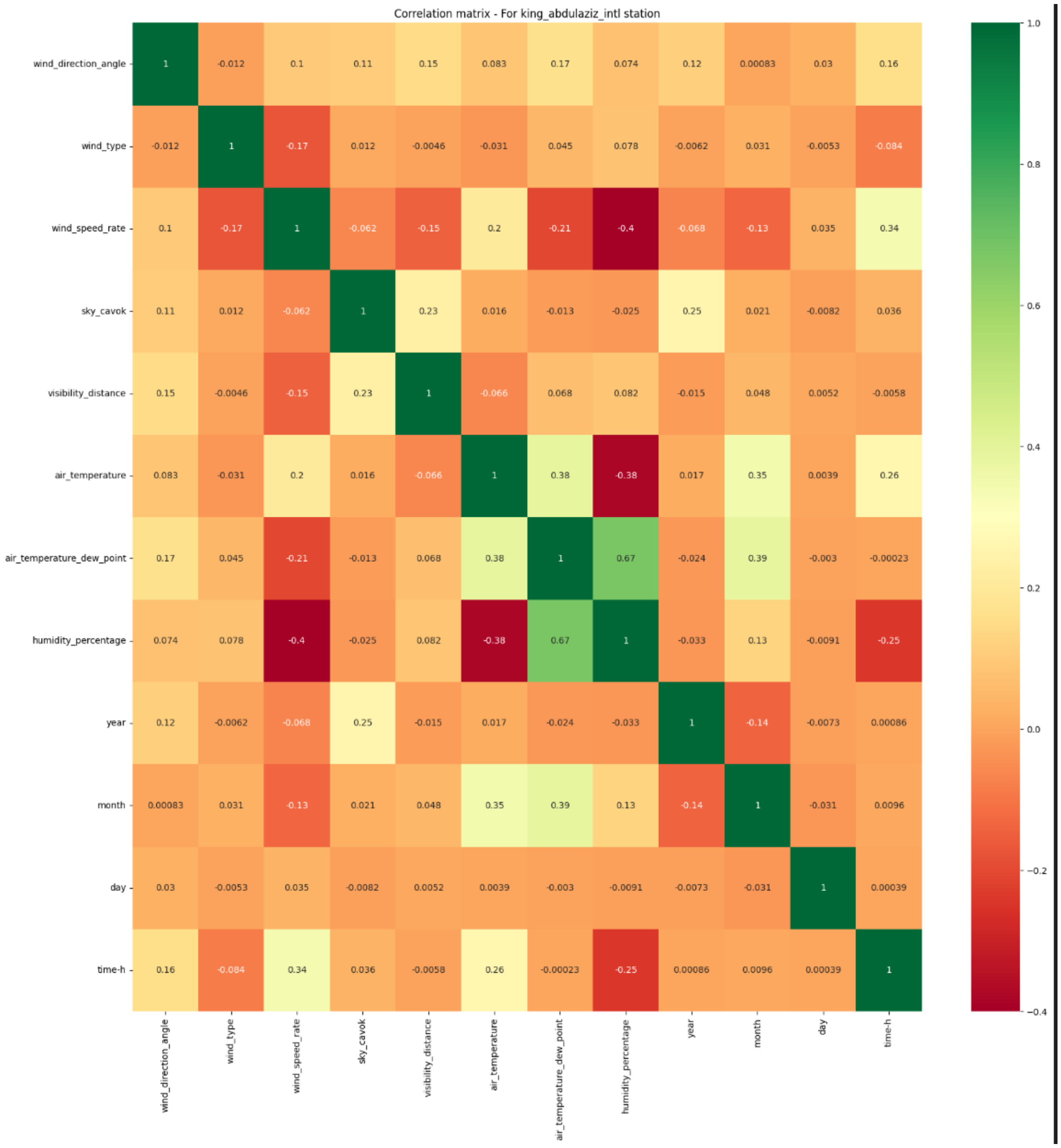
We also have done a correlation matrix for the entire dataset including the three stations together. to show how strong is the relation between the features.

We have done four correlation matrixes, one as we mentioned for the entire data, and we have done each station individually.



- There is a 0.26 relation between wind direction and air temperature dew point. So, we can say, wind direction affects the air temperature dew point.
- Wind speed shows a high relation with humidity. So, we can say humidity affects wind speed.
- Temperature and humidity not showing a considerable relation with wind direction.
- Also, wind speed and direction do not affect visibility.

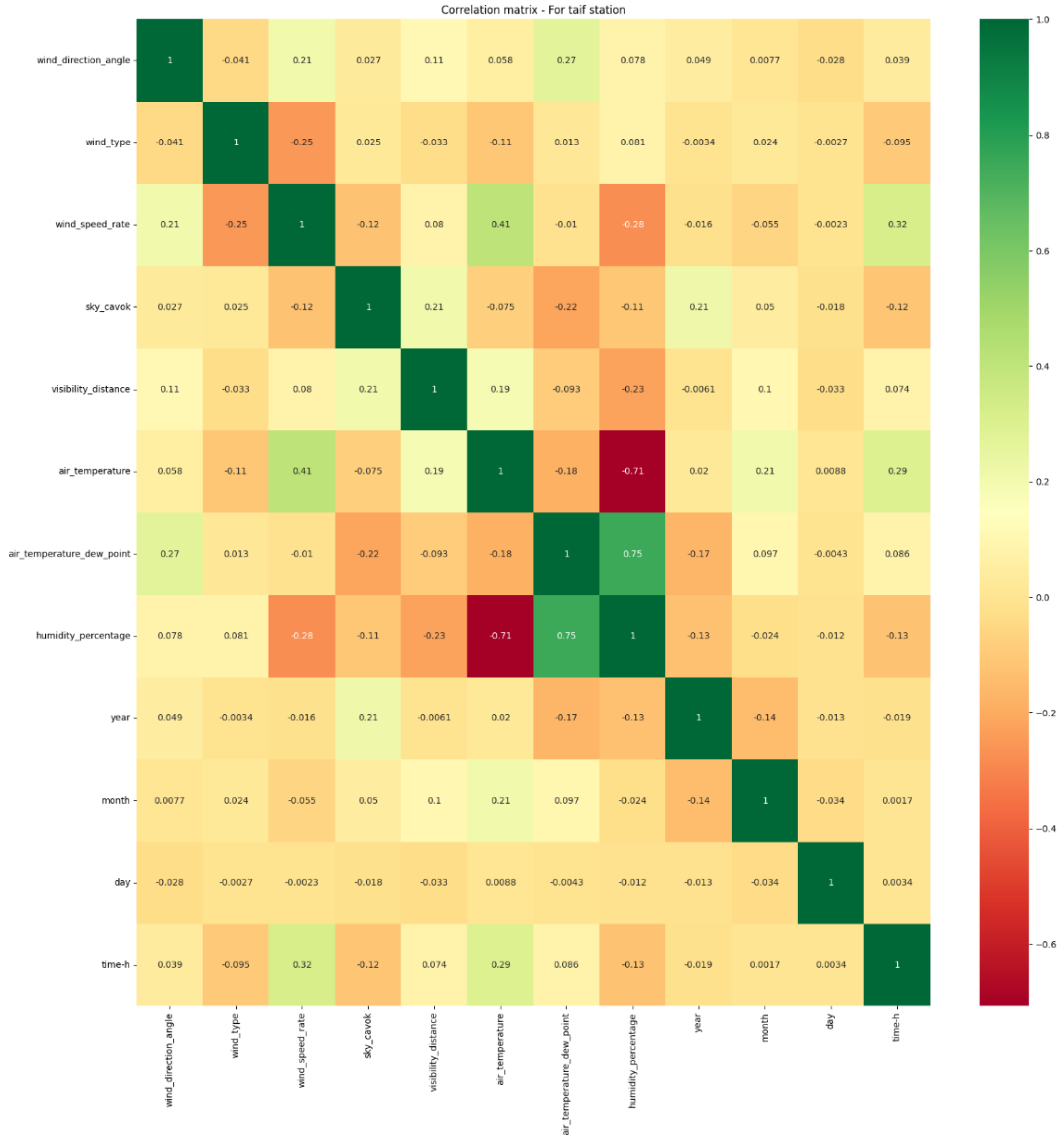
King Abdulaziz International Airport Station correlation matrix:



	wind_direction_angle	wind_type	wind_speed_rate	sky_cavok	visibility_distance	air_temperature	air_temperature_dew_point	humidity_percentage	year	month	day	time-h
wind_direction_angle	1	-0.012	0.1	0.11	0.15	0.083	0.17	0.074	0.12	0.00083	0.03	0.16
wind_type	-0.012	1	-0.17	0.012	-0.0046	-0.031	0.045	0.078	-0.0062	0.031	-0.0053	-0.084
wind_speed_rate	0.1	-0.17	1	-0.062	-0.15	0.2	-0.21	-0.4	-0.068	-0.13	0.035	0.34
sky_cavok	0.11	0.012	-0.062	1	0.23	0.016	-0.013	-0.025	0.25	0.021	-0.0082	0.036
visibility_distance	0.15	-0.0046	-0.15	0.23	1	-0.066	0.068	0.082	-0.015	0.048	0.0052	-0.0058
air_temperature	0.083	-0.031	0.2	0.016	-0.066	1	0.38	-0.38	0.017	0.35	0.0039	0.26
air_temperature_dew_point	0.17	0.045	-0.21	-0.013	0.068	0.38	1	0.67	-0.024	0.39	-0.003	-0.00023
humidity_percentage	0.074	0.078	-0.4	-0.025	0.082	-0.38	0.67	1	-0.033	0.13	-0.0091	-0.25
year	0.12	-0.0062	-0.068	0.25	-0.015	0.017	-0.024	-0.033	1	-0.14	-0.0073	0.00086
month	0.00083	0.031	-0.13	0.021	0.048	0.35	0.39	0.13	-0.14	1	-0.031	0.0096
day	0.03	-0.0053	0.035	-0.0082	0.0052	0.0039	-0.003	-0.0091	-0.0073	-0.031	1	0.00039
time-h	0.16	-0.084	0.34	0.036	-0.0058	0.26	-0.00023	-0.25	0.00086	0.0096	0.00039	1

- wind speed rate and the air_temp_dew_point have weak negative relation with (-.21)
- wind speed rate and visibility have a very weak negative relation with (-0.15)
- air_temperature and the wind speed rate have a weak positive relation with (0.2)

Taif Station correlation matrix:

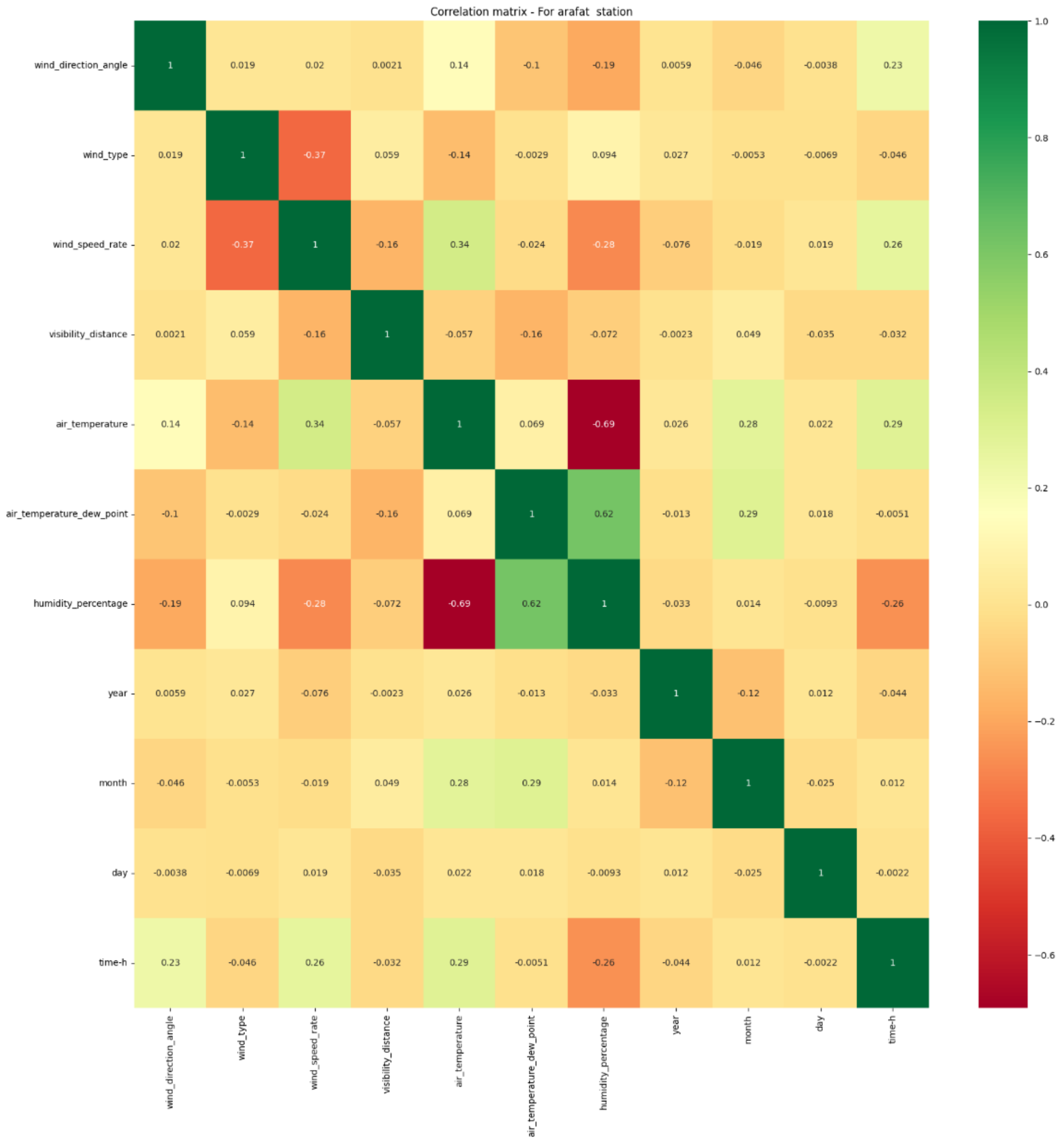


Correlation matrix - For taif station

	wind_direction_angle	wind_type	wind_speed_rate	sky_cavok	visibility_distance	air_temperature	air_temperature_dew_point	humidity_percentage	year	month	day	time-h
wind_direction_angle	1	-0.041	0.21	0.027	0.11	0.058	0.27	0.078	0.049	0.0077	-0.028	0.039
wind_type	-0.041	1	-0.25	0.025	-0.033	-0.11	0.013	0.081	-0.0034	0.024	-0.0027	-0.095
wind_speed_rate	0.21	-0.25	1	-0.12	0.08	0.41	-0.01	-0.28	-0.016	-0.055	-0.0023	0.32
sky_cavok	0.027	0.025	-0.12	1	0.21	-0.075	-0.22	-0.11	0.21	0.05	-0.018	-0.12
visibility_distance	0.11	-0.033	0.08	0.21	1	0.19	-0.093	-0.23	-0.0061	0.1	-0.033	0.074
air_temperature	0.058	-0.11	0.41	-0.075	0.19	1	-0.18	-0.71	0.02	0.21	0.0088	0.29
air_temperature_dew_point	0.27	0.013	-0.01	-0.22	-0.093	-0.18	1	0.75	-0.17	0.097	-0.0043	0.086
humidity_percentage	0.078	0.081	-0.28	-0.11	-0.23	-0.71	0.75	1	-0.13	-0.024	-0.012	-0.13
year	0.049	-0.0034	-0.016	0.21	-0.0061	0.02	-0.17	-0.13	1	-0.14	-0.013	-0.019
month	0.0077	0.024	-0.055	0.05	0.1	0.21	0.097	-0.024	-0.14	1	-0.034	0.0017
day	-0.028	-0.0027	-0.0023	-0.018	-0.033	0.0088	-0.0043	-0.012	-0.013	-0.034	1	0.0034
time-h	0.039	-0.095	0.32	-0.12	0.074	0.29	0.086	-0.13	-0.019	0.0017	0.0034	1

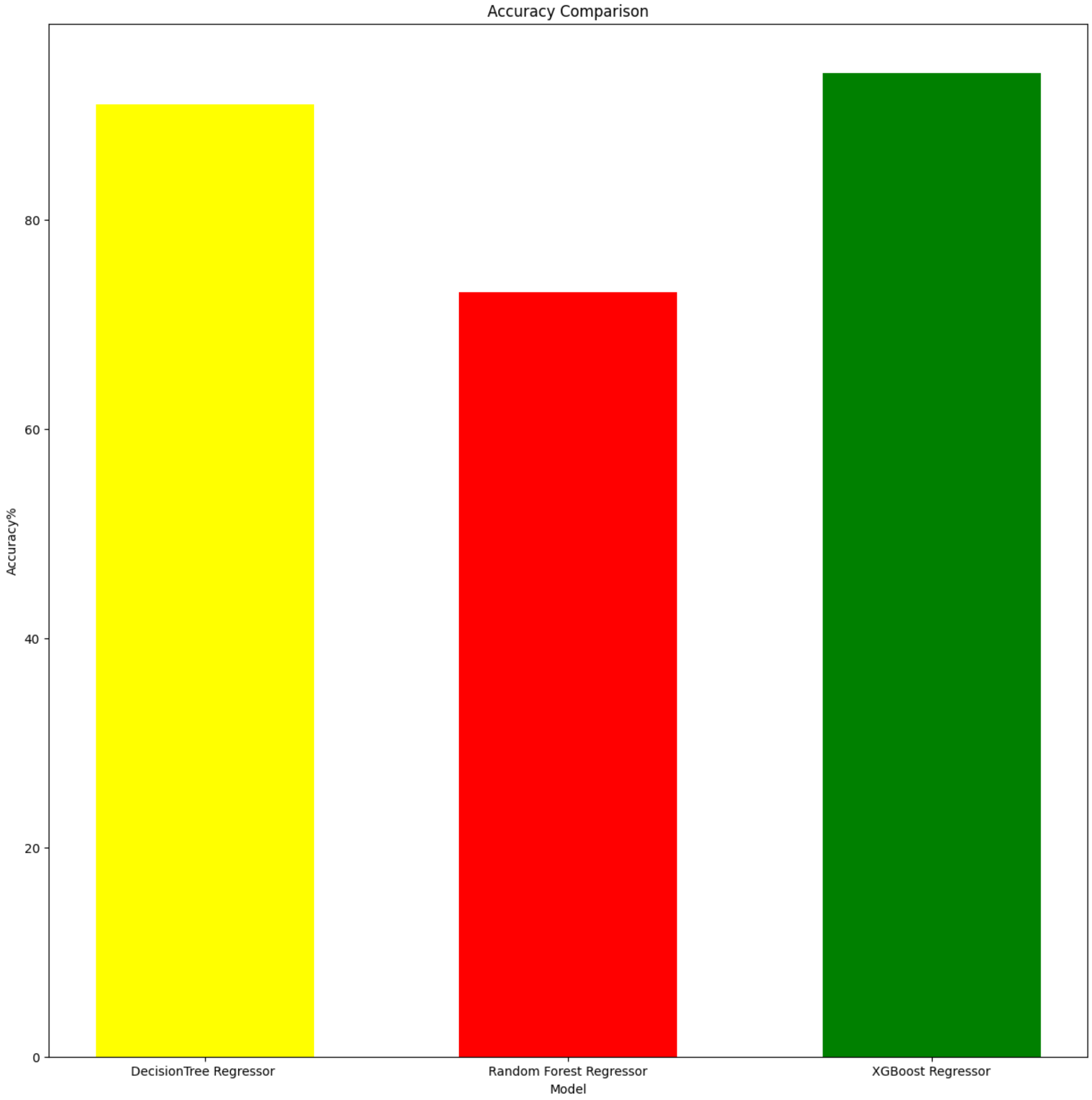
- wind direction angle and air_temp_dew_point have a weak positive relation with (0.27)
- wind speed rate and wind angle have a weak positive relation with (0.21)
- wind angle and air_temp have a very weak positive relation with (0.058)
- wind angle with visibility has a very weak positive relation with (0.11)
- wind speed rate and air_temp have a significant positive relation with (0.41)

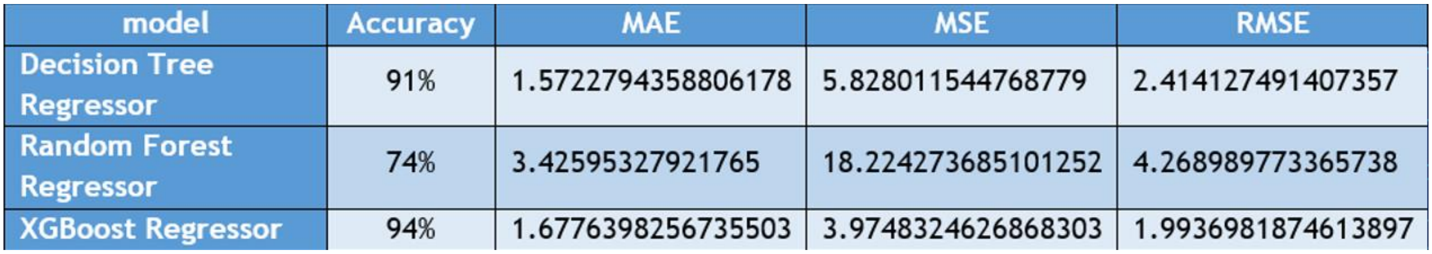
Arafat Station (Makkah) correlation matrix:



- wind speed and air_temp have a weak positive relation with (0.34)
- wind speed and visibility have a very weak negative relation with (-0.16)

We tested five different algorithms with our data to find out which is the best model that fits perfectly with our dataset. And as you can see XGBoost model has to represent the best accuracy with 94%, on other hand the lowest was Random Forest Regression shows 74%.





model	Accuracy	MAE	MSE	RMSE
Decision Tree Regressor	91%	1.5722794358806178	5.828011544768779	2.414127491407357
Random Forest Regressor	74%	3.42595327921765	18.224273685101252	4.268989773365738
XGBoost Regressor	94%	1.6776398256735503	3.9748324626868303	1.9936981874613897

Prediction 6 days ahead for the tree stations, King Abdulaziz, Taif, Arafat.
The prediction will include the following:

Wind speed, wind direction, temperature, visibility.

1-King Abdulaziz station:

Days	Wind speed	wind direction	temperature	visibility
1	14.823844	314.61008	24.468542	9.907873
2	18.44907	318.17523	24.04045	9.314305
3	14.262039	281.5005	23.261915	9.469678
4	10.573023	286.68808	24.052761	9.902076
5	22.147772	297.3704	22.951542	9.761612
6	23.95867	269.82962	22.280882	9.76048

2- Taif station:

Days	Wind speed	wind direction	temperature	visibility
1	11.303469	274.9912	18.368938	10.043708
2	12.735363	256.32455	18.726875	9.983716
3	15.091217	226.82106	18.73921	10.132367
4	17.655487	242.13834	19.090668	9.83247
5	21.212723	256.6115	17.671503	9.73341
6	17.95798	243.48634	14.514509	9.822519

3-Arafat station:

Days	Wind speed	wind direction	temperature	visibility
1	9.43433	282.80582	28.571476	10.063606
2	9.113719	295.32495	27.957281	10.198971
3	6.9383607	234.54204	26.13783	9.720888
4	11.326666	252.57162	25.276468	9.860897
5	7.928068	268.4261	21.974503	9.70511
6	9.339793	232.74173	21.878357	9.967508

6.4 Discussion

We were able to compare the features much better after preprocessing the data and splitting up the three stations. Our goal through this process was to figure out which features had strong relationships, and which did not. In addition, we wanted to find out what the most suitable model is for training and testing our data. We figured out that wind speed does not have a significant effect on visibility. However, we also found that wind speed has a strong influence on visibility in different months.

6.5 Conclusion

Most of our dataset’s frequency belonged to King Abdulaziz International Airport station, followed by Taif's station, and then Arafat's station. After we tested the models, we found out that the XGBoost model had the highest accuracy, while the Random Forest Regression model had the lowest accuracy. We observed that the relationship between humidity and wind speed rate at KAIA station is very strong, whereas, at Taif and Arafat stations, it is not the same. Thus, we can conclude that humidity affects wind speed. At Taif's station, the wind speed rate was the highest when compared to KAIA & Arafat stations.

Chapter 7: Conclusion & Future Work

7.1 Introduction

In this chapter, we will discuss our final thoughts and conclusions from this project. We will also discuss the challenges and limitations we faced and how we solved them, and possible ideas to expand on the project in the section on the future.

7.2 Conclusion

The weather dataset was significant. So, we were intrigued by the subject and wanted to study and understand it more before building the model. We did so by doing multiple exploratory data analyses on each feature. Then we started cleaning the dataset to ensure that it was ready for predictive analysis because we learned in our studies that "Clean input = Clean output," so we had to ensure that the data was ready and clean. As we started building the classification models, we faced multiple difficulties that will be discussed in detail in the next section. However, in the end, we managed to find an optimal accuracy of 94% using XGBoost. However, for the predicting part, LSTM proved to be the optimal algorithm.

7.3 Difficulties & Limitations

During the model-building phase, we encountered many challenges and limitations with the results. After finding out that the Linear Regression model was so low in accuracy, we couldn’t keep it with the project. The same goes for the Support Vector Regressor model it was even lower than the linear regression model. So, we had to remove both and continue with the rest of the algorithms.

7.4 Future Work

We are interested in working on similar ideas and experimenting with new stations, such as the Riyadh station. Alternatively, we could establish a project that can be provided to institutions and businesses that require wind speed and direction angle information to help them achieve their goals.

References:

-Bobbitt, Z. (2020) *How to read a correlation matrix*, *Statology*. Zach. Available at: https://www.statology.org/how-to-read-a-correlation-matrix/ (Accessed: February 13, 2023).

-Cherry, K. (2022) *The role of correlations in psychology research*, *Verywell Mind*. Verywell Mind. Available at: https://www.verywellmind.com/what-is-correlation-2794986 (Accessed: February 13, 2023).

-Ali, Z. (2020) *Wind direction and speed prediction using machine learning in Python*, *CodeSpeedy*. CodeSpeedy. Available at: https://www.codespeedy.com/wind-direction-and-speed-prediction-using-machine-learning-in-python/ (Accessed: February 13, 2023).

-Kumarl, A. (2022) *Correlation Concepts, Matrix & Heatmap using Seaborn*, *Data Analytics*. Data Analytics. Available at: https://vitalflux.com/correlation-heatmap-with-seaborn-pandas/ (Accessed: February 13, 2023).

-Jeet, M. (2022) *Python strings decode() method*, *GeeksforGeeks*. GeeksforGeeks. Available at: https://www.geeksforgeeks.org/python-strings-decode-method/ (Accessed: February 13, 2023).

-Brary, S. (2016) *Meteorological terminal air report (METAR)*, *SKYbrary Aviation Safety*. SKYbrary Aviation Safety. Available at: https://www.skybrary.aero/articles/meteorological-terminal-air-report-metar (Accessed: February 13, 2023).

-Singh, P. and Singh, A. (2021) *Confusion matrix*, *Confusion Matrix - an overview | ScienceDirect Topics*. Confusion Matrix. Available at: https://www.sciencedirect.com/topics/engineering/confusion-matrix (Accessed: February 13, 2023).

-Shop, D. (2022) *What does inplace mean in pandas?*, *GeeksforGeeks*. GeeksforGeeks. Available at: https://www.geeksforgeeks.org/what-does-inplace-mean-in-pandas/ (Accessed: February 13, 2023).

-Campbell, S. (2023) *Scipy in python tutorial: What is, Library, Function & Examples*, *Guru99*. Guru99. Available at: https://www.guru99.com/scipy-tutorial.html (Accessed: February 13, 2023).

-M, R. (2022) *What is Matplotlib in python? how to use it for plotting?*, *ActiveState*. ActiveState. Available at: https://www.activestate.com/resources/quick-reads/what-is-matplotlib-in-python-how-to-use-it-for-plotting/ (Accessed: February 13, 2023).

- Wiksom, M. (2022) *Statistical Data Visualization#*, *seaborn*. seaborn. Available at: https://seaborn.pydata.org/ (Accessed: February 13, 2023).

-*What is a decision tree* (2023) *IBM*. IBM. Available at: https://www.ibm.com/sa-en/topics/decision-trees (Accessed: February 13, 2023).

- Chaya (2022) *Random Forest regression*, *Medium*. Level Up Coding. Available at: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84 (Accessed: February 13, 2023).

- Brownlee, J. (2021) *XGBoost for regression*, *MachineLearningMastery.com*. MachineLearningMastery.com. Available at: https://machinelearningmastery.com/xgboost-for-regression/ (Accessed: February 13, 2023).

- Raj, A. (2020) *Unlocking the true power of support vector regression*, *Medium*. Towards Data Science. Available at: https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0 (Accessed: February 13, 2023).

- Panik, M. (2021) *Mean squared error: Definition and example*, *Statistics How To*. Available at: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-squared-error/ (Accessed: February 13, 2023).

- M, P. (2022) *Evaluation metric for regression models*, *Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/#:~:text=Mean%20Absolute%20Error%20(MAE),-Mean%20absolute%20error&text=It%20is%20calculated%20by%20taking,arithmetic%20average%20of%20absolute%20errors. (Accessed: February 13, 2023).

- Barnston, A. (2022) *RMSE: Root mean square error*, *Statistics How To*. Available at: https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/ (Accessed: February 13, 2023).

- Kanade, V. (2022) *What is linear regression? types, equation, examples, and best practices for 2022*, *Spiceworks*. Available at: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/ (Accessed: February 13, 2023).