WEATHER DATA ANALYSIS

18CSE394T – Business Intelligence & Analytics

Mini Project Report

Submitted by

RISHIRAJ SAHA [Reg. No.: RA2011027010015]
B.Tech. CSE – Big Data Analytics

SCHOOL OF COMPUTING
COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Under Section 3 of UGC Act, 1956)
S.R.M. NAGAR, KATTANKULATHUR – 603 203
CHENGALPATTU DISTRICT

November 2023

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)

COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Under Section 3 of UGC Act, 1956)
S.R.M. NAGAR, KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that Mini project report titled **WEATHER DATA ANALYSIS** is the bonafide work of Reg.No RA2011027010015 Name Rishiraj Saha who carried out the minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. A Shantini

Assistant Professor

(GUIDE)

SIGNATURE

Dr. M LAKSHMI

(HEAD OF THE DEPARTMENT)

TABLE OF CONTENTS

1. Problem Statement

In this weather data analysis project, we aim to investigate and derive meaningful insights from a diverse set of meteorological data. Focusing on historical weather patterns, temperature variations, precipitation levels, and atmospheric conditions, our objective is to identify trends, anomalies, and potential correlations within the dataset. By employing advanced statistical and machine learning techniques, we seek to enhance our understanding of local and global climate phenomena. The outcomes of this analysis will not only contribute to scientific research but also have practical applications in areas such as agriculture, disaster preparedness, and urban planning. Through meticulous data exploration, modeling, and visualization, this project aims to unravel the intricate dynamics of weather systems for comprehensive and actionable conclusions.

2. Methodology / Procedure/ Algorithm

Methodology for the Weather Data Analysis:

The first step involves comprehensive data collection from reliable meteorological sources, encompassing variables such as temperature, precipitation, humidity, wind speed, and atmospheric pressure. A meticulous approach is taken to address data integrity issues, including the handling of missing values, outliers, and inconsistencies.

Following data collection, a thorough Exploratory Data Analysis (EDA) is conducted to gain insights into data distributions, correlations, and initial patterns. This includes the visualization of temporal trends, seasonal variations, and geographical differences to inform subsequent analytical steps.

Feature engineering is then employed to derive relevant features, such as monthly averages, seasonal indicators, and anomaly scores. Additionally, the incorporation of external factors, such as geographical features or socio-economic data, is considered where applicable.

Statistical analysis is a key component of the methodology, involving the application of methods to identify significant trends, correlations, and anomalies. Hypothesis testing is performed to validate findings and assess the statistical significance of observed patterns.

Machine learning models, including regression and time series analysis, are implemented for predictive modeling. These models are trained using historical data to forecast future weather conditions, enhancing the project's predictive capabilities.

Spatial analysis comes into play through the utilization of geospatial techniques to analyze and visualize spatial patterns in weather data. This step allows for the exploration of regional variations and an assessment of the impact of geography on weather patterns.

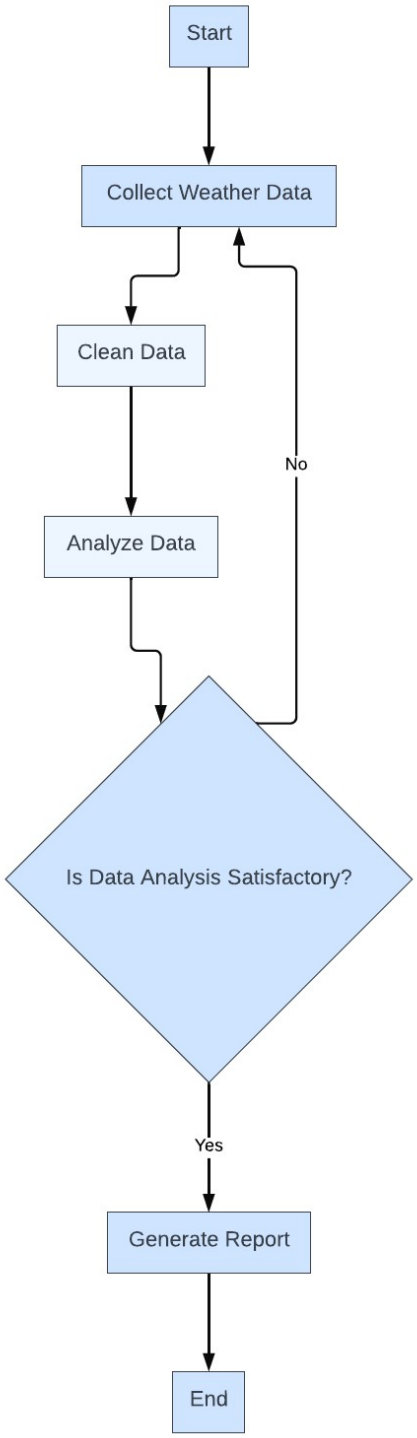
Validation and model tuning follow, where models are rigorously validated using appropriate metrics and cross-validation techniques. Fine-tuning is then executed based on validation results to improve overall accuracy and reliability.

Interpretation and reporting involve contextualizing findings within the realm of meteorological knowledge and real-world implications. Detailed reports, inclusive of visualizations, key insights, and recommendations, are prepared for stakeholders.

Effective communication is maintained throughout the project by engaging with domain experts, stakeholders, and the community. This iterative process ensures continuous refinement of analyses and a clear articulation of methodologies, limitations, and implications.

Finally, meticulous documentation captures the entire process, including data preprocessing steps, analysis code, and model specifications. This commitment to documentation aims to ensure transparency and reproducibility for future reference and validation.

3. Flow chart


Start
Collect Weather Data
Clean Data
Analyze Data
Is Data Analysis Satisfactory?
No
Yes
Generate Report
End

4. Coding (Python)

This section contains the code for data preparation where dataset is explored cleaned, new features are added if required all other processes which will make data much better to handle and create dashboards.

```
[6]  data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8784 entries, 0 to 8783
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Date/Time         8784 non-null   object 
 1   Temp_C            8784 non-null   float64
 2   Dew Point Temp_C  8784 non-null   float64
 3   Rel Hum_%         8784 non-null   int64  
 4   Wind Speed_km/h   8784 non-null   int64  
 5   Visibility_km     8784 non-null   float64
 6   Press_kPa         8784 non-null   float64
 7   Weather           8784 non-null   object 
dtypes: float64(4), int64(2), object(2)
memory usage: 549.1+ KB
```

```
[11] data['Weather'].unique()
```

```
array(['Fog', 'Freezing Drizzle,Fog', 'Mostly Cloudy', 'Cloudy', 'Rain',
       'Rain Showers', 'Mainly Clear', 'Snow Showers', 'Snow', 'Clear',
       'Freezing Rain,Fog', 'Freezing Rain', 'Freezing Drizzle',
       'Rain,Snow', 'Moderate Snow', 'Freezing Drizzle,Snow',
       'Freezing Rain,Snow Grains', 'Snow,Blowing Snow', 'Freezing Fog',
       'Haze', 'Rain,Fog', 'Drizzle,Fog', 'Drizzle',
       'Freezing Drizzle,Haze', 'Freezing Rain,Haze', 'Snow,Haze',
       'Snow,Fog', 'Snow,Ice Pellets', 'Rain,Haze', 'Thunderstorms,Rain',
       'Thunderstorms,Rain Showers', 'Thunderstorms,Heavy Rain Showers',
       'Thunderstorms,Rain Showers,Fog', 'Thunderstorms',
       'Thunderstorms,Rain,Fog',
       'Thunderstorms,Moderate Rain Showers,Fog', 'Rain Showers,Fog',
       'Rain Showers,Snow Showers', 'Snow Pellets', 'Rain,Snow,Fog',
       'Moderate Rain,Fog', 'Freezing Rain,Ice Pellets,Fog',
       'Drizzle,Ice Pellets,Fog', 'Drizzle,Snow', 'Rain,Ice Pellets',
       'Drizzle,Snow,Fog', 'Rain,Snow Grains', 'Rain,Snow,Ice Pellets',
       'Snow Showers,Fog', 'Moderate Snow,Blowing Snow'], dtype=object)
```

Below I have checked for data entries which have clear weather and wind visibility below 40.

```
#Checking for entries with weather clear and visibility greater than 40
data[(data['Weather'] == 'Clear') | (data['Visibility_km'] > 40)]
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
106	1/5/2012 10:00	-6.0	-10.0	73	17	48.3	100.45	Mainly Clear
107	1/5/2012 11:00	-5.6	-10.2	70	22	48.3	100.41	Mainly Clear
108	1/5/2012 12:00	-4.7	-9.6	69	20	48.3	100.38	Mainly Clear
109	1/5/2012 13:00	-4.4	-9.7	66	26	48.3	100.40	Mainly Clear
...	...	...	...	...	...	...	...	...
8749	12/30/2012 13:00	-12.4	-16.2	73	37	48.3	100.92	Mostly Cloudy
8750	12/30/2012 14:00	-11.8	-16.1	70	37	48.3	100.96	Mainly Clear
8751	12/30/2012 15:00	-11.3	-15.6	70	32	48.3	101.05	Mainly Clear
8752	12/30/2012 16:00	-11.4	-15.5	72	26	48.3	101.15	Mainly Clear
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Clear

3027 rows × 8 columns

Then I have checked which have at least snow under weather and only snow under weather.

```
#Checking which of contains atleast snow under weather
data[data['Weather'].str.contains('Snow')].tail(50)
```

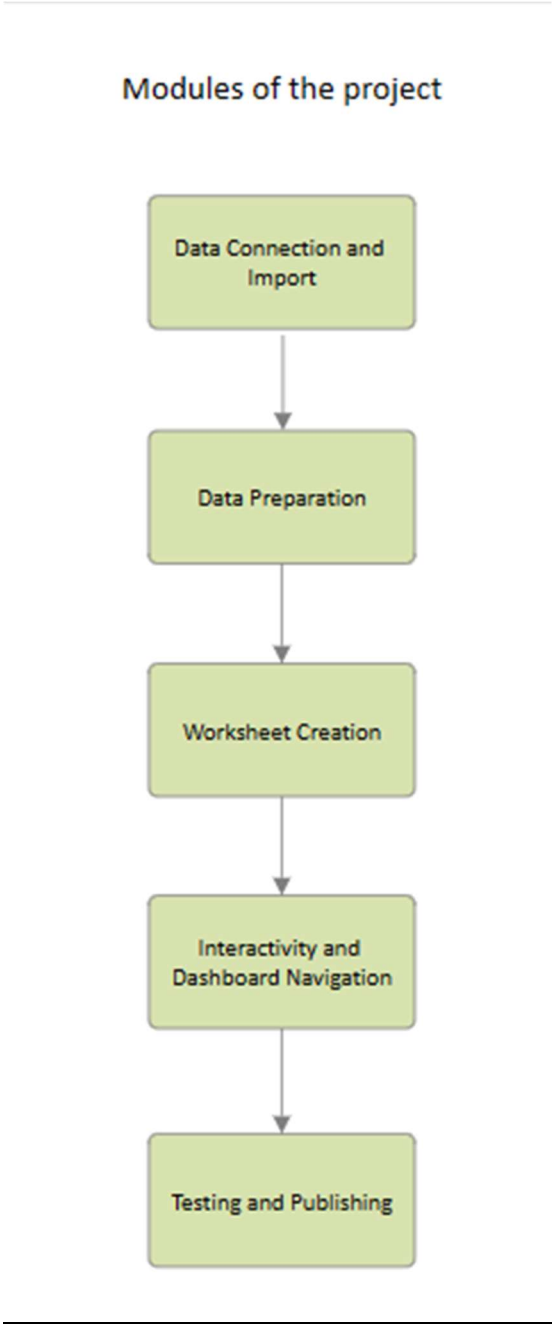
	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
8680	12/27/2012 16:00	-4.5	-6.2	88	37	2.0	100.44	Snow,Blowing Snow
8681	12/27/2012 17:00	-4.2	-5.9	88	32	3.2	100.47	Snow,Blowing Snow
8682	12/27/2012 18:00	-4.0	-5.7	88	28	8.0	100.49	Snow,Blowing Snow
8683	12/27/2012 19:00	-3.9	-5.6	88	26	9.7	100.52	Snow,Blowing Snow
8684	12/27/2012 20:00	-3.7	-5.3	89	37	16.1	100.58	Snow
8685	12/27/2012 21:00	-3.7	-4.8	92	24	4.8	100.62	Freezing Drizzle,Snow
8686	12/27/2012 22:00	-3.8	-4.6	94	20	4.8	100.65	Freezing Drizzle,Snow
8687	12/27/2012 23:00	-4.0	-5.6	89	24	9.7	100.70	Snow
8688	12/28/2012 0:00	-4.2	-5.7	89	19	8.0	100.78	Freezing Drizzle,Snow
8689	12/28/2012 1:00	-4.4	-6.6	85	15	6.4	100.83	Freezing Drizzle,Snow
8690	12/28/2012 2:00	-4.3	-6.3	86	11	12.9	100.93	Freezing Drizzle,Snow
8691	12/28/2012 3:00	-4.6	-5.9	91	13	4.0	101.01	Snow
8692	12/28/2012 4:00	-4.9	-5.9	93	9	9.7	101.00	Snow
8723	12/29/2012 11:00	-10.9	-12.2	90	7	6.4	101.09	Snow Showers,Fog
8724	12/29/2012 12:00	-10.5	-11.6	92	11	8.0	100.93	Snow Showers,Fog
8725	12/29/2012 13:00	-10.0	-11.1	92	22	9.7	100.63	Snow Showers,Fog

```
#Checking for entries with snow weather
data[data['Weather']=='Snow']
```

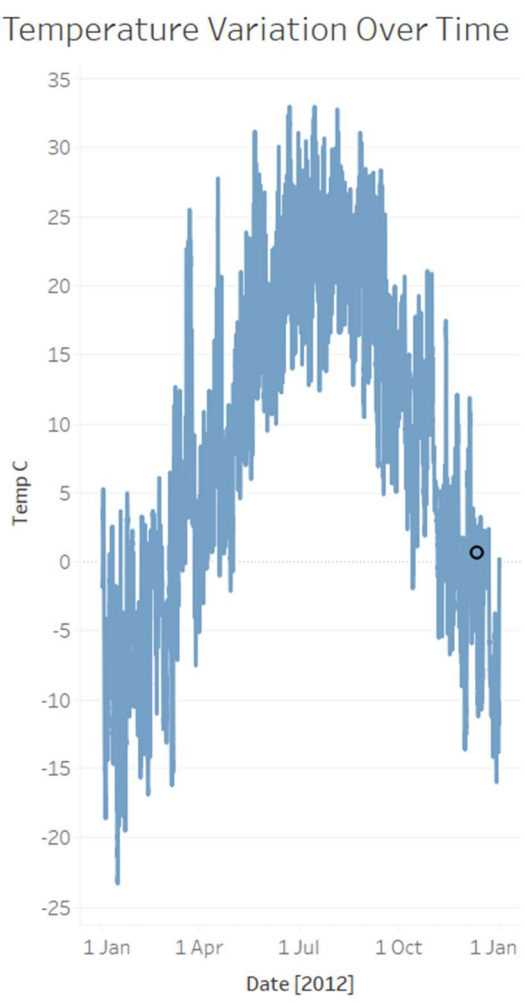
	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
55	1/3/2012 7:00	-14.0	-19.5	63	19	25.0	100.95	Snow
84	1/4/2012 12:00	-13.7	-21.7	51	11	24.1	101.25	Snow
86	1/4/2012 14:00	-11.3	-19.0	53	7	19.3	100.97	Snow
87	1/4/2012 15:00	-10.2	-16.3	61	11	9.7	100.89	Snow
88	1/4/2012 16:00	-9.4	-15.5	61	13	19.3	100.79	Snow
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/31/2012 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

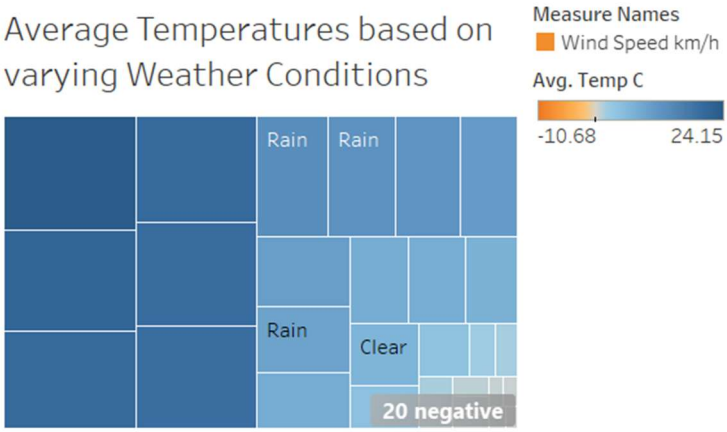
390 rows × 8 columns

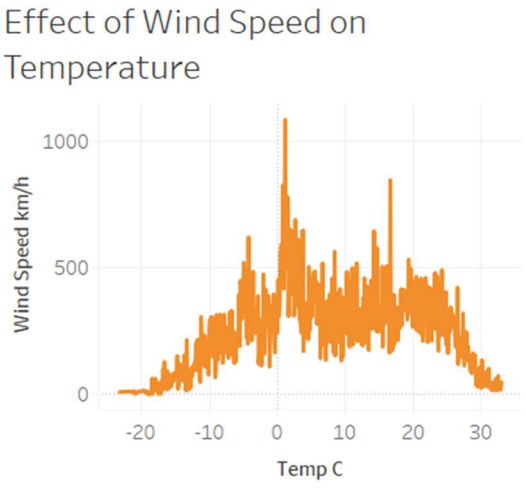
5. Modules of the proposed work



6. Results/Screenshots







7. Conclusion

In summary, this weather data analysis project systematically explored meteorological data, from collection to interpretation. Leveraging advanced statistical methods and machine learning models, it uncovered temporal, spatial, and seasonal patterns. The validated models, tuned for accuracy, enable reliable weather forecasting. Collaborative engagement with stakeholders enriched the analysis, ensuring real-world relevance. The results, communicated through comprehensive reports, provide actionable insights for sectors like agriculture and disaster preparedness. This project not only enhances our understanding of weather dynamics but also offers practical applications, emphasizing the importance of data-driven decision-making in the face of evolving environmental conditions.

8. References

1. **Tableau Desktop Documentation on Official Website**
2. **W3 Schools – Data Science and Python**
3. **Youtube**