INTRODUCTION

Title: "Analysis and Visualization of Hourly Weather Data for NYC Airports (2013)"

Student Name: Vinay Srinivasan

Student Number: "s225447445**"**

Email: "s225447445@deakin.edu.au; s.vinay.777@gmail.com; srinivasan_vinay@yahoo.com;"

Public Repository: https://github.com/SriVinayaka/power-bI-SIG-731-Data-Wrangling-Task-6D.git

https://app.powerbi.com/groups/me/reports/ee09841c-c7c7-401c-9a40-6063c08eae2b/ea122867d30066ad99b0?experience=power-bi

https://app.powerbi.com/groups/me/insights/9be0feed-3ed2-4217-962f-a418ec8679a8?insightsSource=Desktop&experience=power-bi

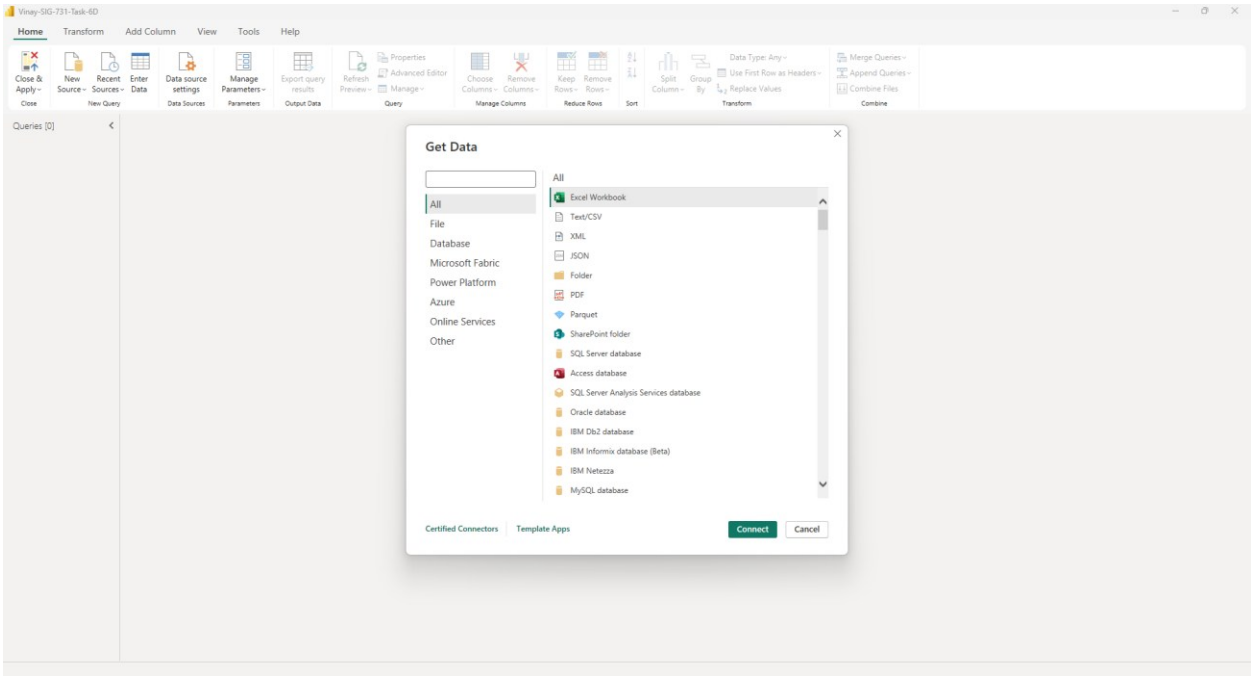
01. Data Source

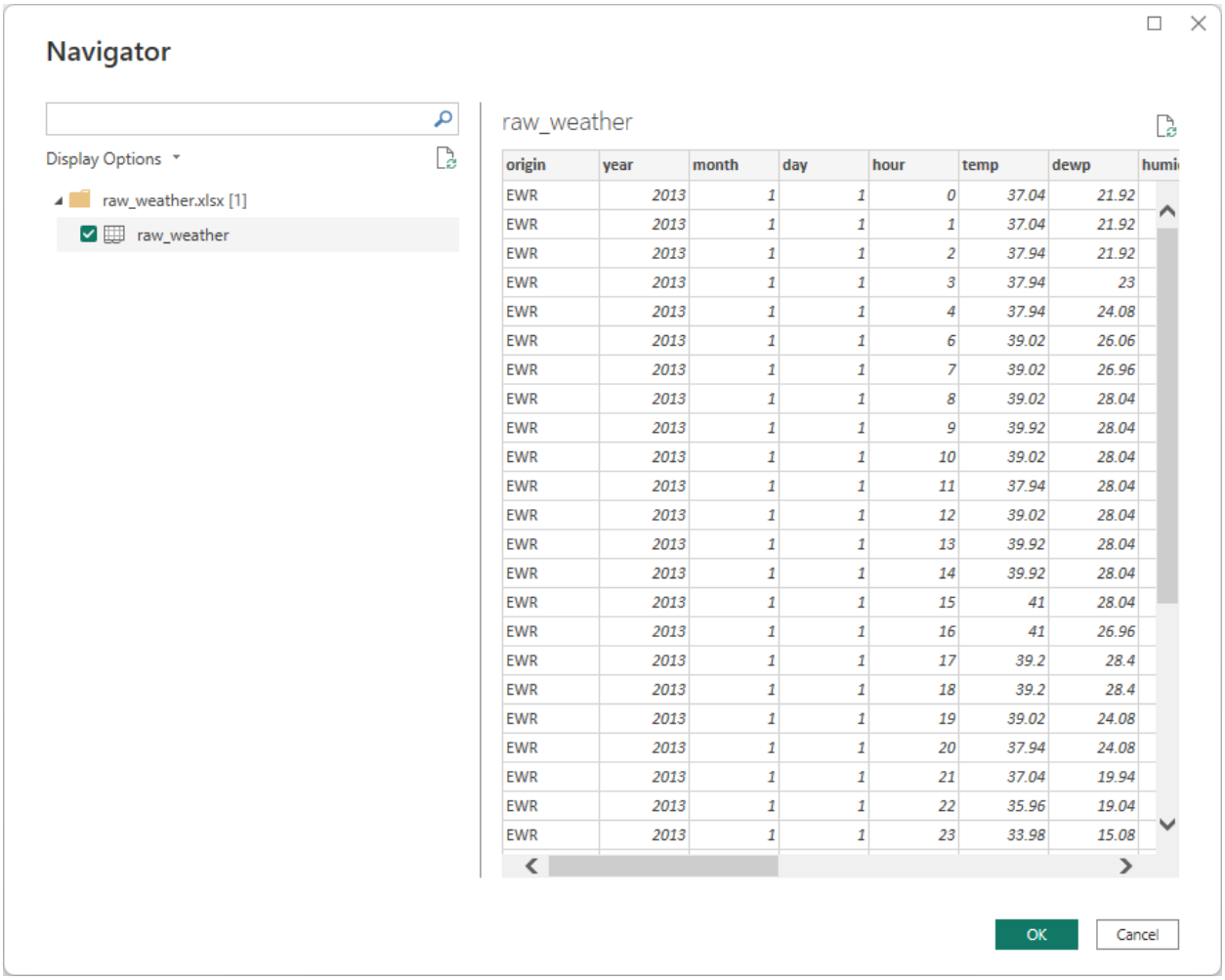
- Dataset: nycflights13_weather.csv.gz
- Extracted file used: weather.csv
- Data description: Hourly meteorological observations for **LGA, JFK, EWR** during 2013
- Convert the weather.csv to raw_weather.xlsx
- Remove the top 42 rows containing the "R" program markers

02. Load Data into Power BI Desktop

1. Open **Power BI Desktop**
2. **Home -> Get Data -> 'Excel Workbook'**
3. Select 'raw_weather.xlsx'
4. Click **Transform Data** (do NOT load directly)

All processing from now on is done in **Power Query Editor** ✔


Vinay-SIG-731-Task-6D
Home
Transform
Add Column
View
Tools
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Export query results
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Any
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Output Data
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [0]
Get Data
All
File
Database
Microsoft Fabric
Power Platform
Azure
Online Services
Other
All
Excel Workbook
Text/CSV
XML
JSON
Folder
PDF
Parquet
SharePoint folder
SQL Server database
Access database
SQL Server Analysis Services database
Oracle database
IBM Db2 database
IBM Informix database (Beta)
IBM Netezza
MySQL database
Certified Connectors
Template Apps
Connect
Cancel



origin	year	month	day	hour	temp	dewp	humi
EWR	2013	1	1	0	37.04	21.92	
EWR	2013	1	1	1	37.04	21.92	
EWR	2013	1	1	2	37.94	21.92	
EWR	2013	1	1	3	37.94	23	
EWR	2013	1	1	4	37.94	24.08	
EWR	2013	1	1	6	39.02	26.06	
EWR	2013	1	1	7	39.02	26.96	
EWR	2013	1	1	8	39.02	28.04	
EWR	2013	1	1	9	39.92	28.04	
EWR	2013	1	1	10	39.02	28.04	
EWR	2013	1	1	11	37.94	28.04	
EWR	2013	1	1	12	39.02	28.04	
EWR	2013	1	1	13	39.92	28.04	
EWR	2013	1	1	14	39.92	28.04	
EWR	2013	1	1	15	41	28.04	
EWR	2013	1	1	16	41	26.96	
EWR	2013	1	1	17	39.2	28.4	
EWR	2013	1	1	18	39.2	28.4	
EWR	2013	1	1	19	39.02	24.08	
EWR	2013	1	1	20	37.94	24.08	
EWR	2013	1	1	21	37.04	19.94	
EWR	2013	1	1	22	35.96	19.04	
EWR	2013	1	1	23	33.98	15.08	

Question 1: Data Cleaning and Preparation

master-cleaned-data

The raw weather dataset was imported into Power BI and cleaned to ensure consistency and accuracy. Airport codes were standardized, non-numeric wind speed values (NA) were handled appropriately, and wind speed measurements were converted from miles per hour (mph) to meters per second (m/s). A proper date structure was created using year, month, and day fields.

qns001-data

After cleaning, the dataset contained valid numerical wind speed values in SI units and consistent airport identifiers. This ensured the data was suitable for aggregation, comparison across airports, and further statistical analysis.

Question 2: Daily Mean Wind Speed Calculation for LGA

qns002-data

Hourly wind speed observations were aggregated to compute daily mean wind speeds for LGA airport. Aggregation was performed at the query level to avoid duplication and incorrect summarization during visualization.

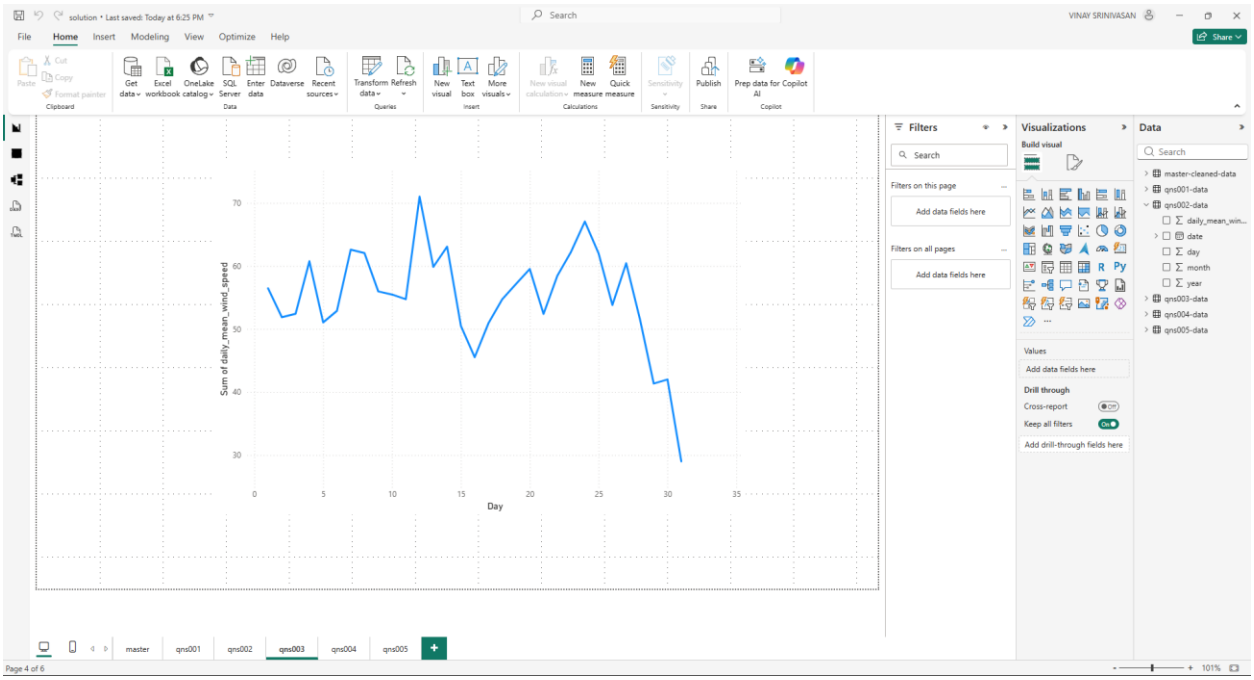
The resulting dataset contained one observation per airport per day. This reduced data granularity while preserving daily wind patterns and ensured accurate trend analysis without aggregation distortion.

Question 3: Daily Wind Speed Trend Analysis for LGA (line graph)

qns003-data

Daily mean wind speeds were visualized over time using line charts to examine temporal trends. Separate trends were maintained for LGA airport to enable direct comparison based on Days

The plots revealed clear temporal variations in wind speed, including seasonal patterns. Differences between airports were observable, demonstrating how wind conditions vary geographically over time.

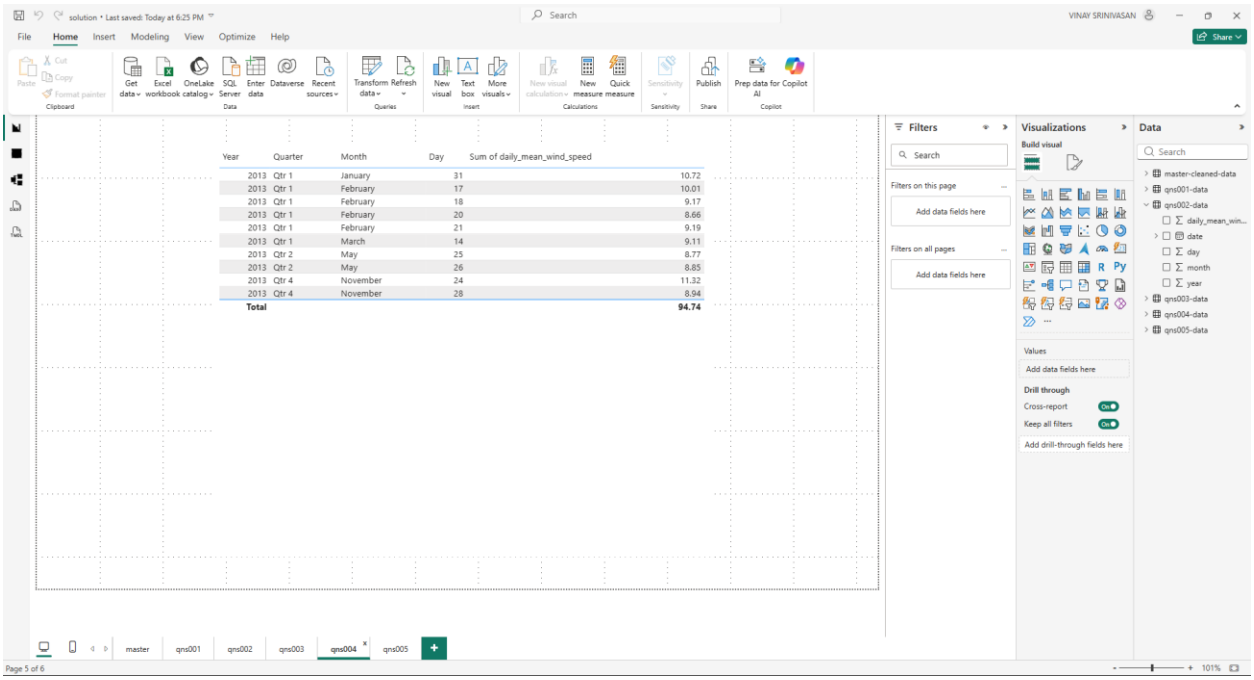


Question 4: Identification of Windiest Days for LGA (Table Format)

qns004-data

Daily mean wind speeds were ranked to identify the top ten windiest days for LGA airport. This analysis focused on extreme wind conditions while maintaining data integrity.

The results highlighted specific dates with unusually high wind speeds at LGA airport. These events represent periods of extreme weather and provide insight into peak wind conditions across that location.

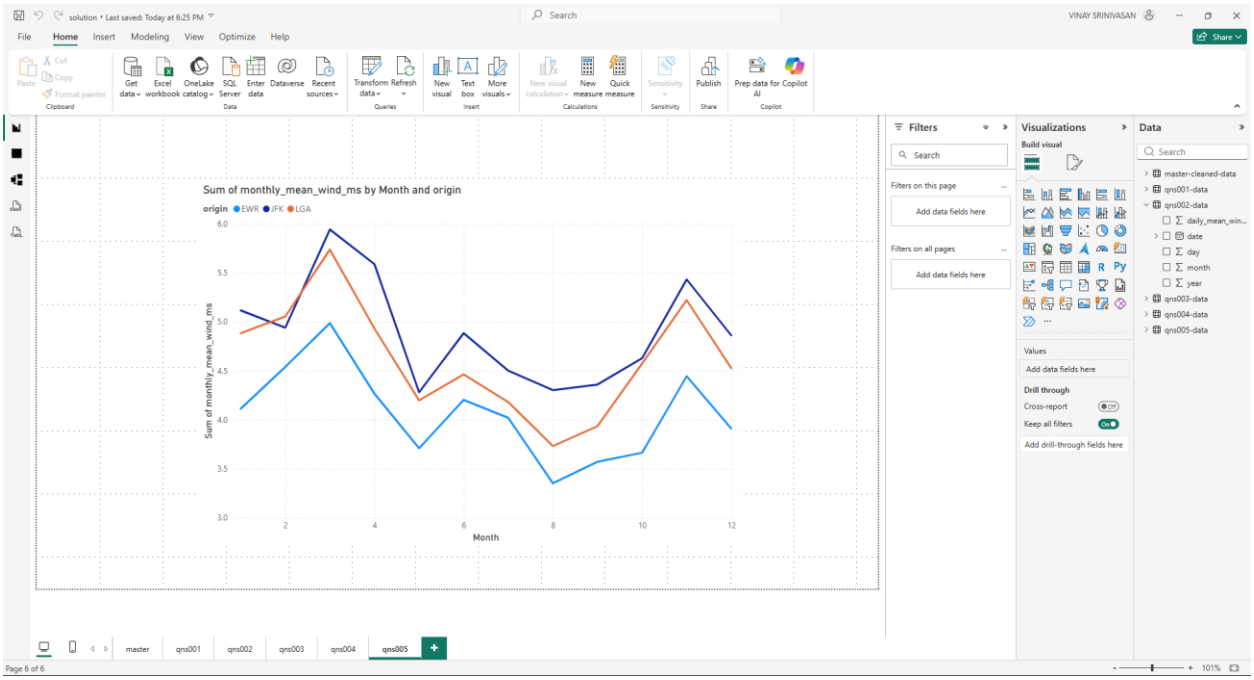


Question 5: Monthly Mean Wind Speed and Outlier Removal (Line Graph for EWR, JFK, LGA)

qns005-data

Hourly wind speed measurements were aggregated to compute monthly mean wind speeds for LGA, JFK, and EWR airports. Statistical outliers were removed using the Interquartile Range (IQR) method, applied independently to each airport.

The cleaned monthly dataset revealed smooth and consistent wind trends across months. Outlier removal reduced the influence of extreme values, allowing clearer comparison of seasonal wind patterns between airports.



Final Dashboard

https://app.powerbi.com/groups/me/reports/f2a65ac2-daf8-4fec-a3e3-fcd4ca5311fe/c0f693e5777160b349ed?experience=power-bi

https://app.powerbi.com/groups/me/insights/9be0feed-3ed2-4217-962f-a418ec8679a8?insightsSource=Desktop&experience=power-bi

All meteorological variables were converted to International System of Units (SI) or SI-derived units using Power Query. Original measurements were preserved, and new columns were created for temperature (°C), wind speed (m/s), precipitation (mm), visibility (km), pressure (hPa), and relative humidity (fraction). This approach ensures scientific consistency and reproducibility

Line Chart

The line chart shows a strong seasonal cycle in monthly mean wind speed, with higher values during winter and lower values in summer. EWR generally experiences the strongest winds, while LGA records the weakest, although all airports follow similar temporal trends

Table Chart

The table chart displays exact monthly mean wind speed values by airport, supporting transparency and confirming the seasonal and inter-airport patterns observed in the graphical visualizations

Conclusion

The step-by-step data wrangling process transformed raw hourly weather observations into clean, aggregated, and analysis-ready datasets. By performing all transformations within Power BI's Advanced Query Editor, aggregation errors were avoided and data integrity was preserved. The resulting analyses provided meaningful insights into daily and monthly wind behavior across major airports. The Final Power-BI Report has been published to:

https://app.powerbi.com/groups/me/reports/f2a65ac2-daf8-4fec-a3e3-fcd4ca5311fe/c0f693e5777160b349ed?experience=power-bi

https://app.powerbi.com/groups/me/insights/9be0feed-3ed2-4217-962f-a418ec8679a8?insightsSource=Desktop&experience=power-bi

https://github.com/SriVinayaka/power-bI-SIG-731-Data-Wrangling-Task-6D.git

Artifacts:

Public Repository: https://github.com/SriVinayaka/power-bI-SIG-731-Data-Wrangling-Task-6D.git

All Artefacts (Power-BI Model Advanced Queries) are stored in "queries" folder and data is stored in "data" folder