Task-5C: Power BI Dashboards

Name: Yatharth Deoly
StudentId: 224207854
EmailId: yatharthdeoly@gamail.com

Introduction

In this case study we will be learning the usage of Microsoft Power BI with help of the Airline on-time data for all flights departing from NYC in 2013. This dataset contains different features i.e.:

- `origin` –> weather station: LGA, JFK or EWR.
- `year`, `month`, `day`, `hour` -> time of recording.
- `temp`, `dewp` –> temperature and dew point in degrees Fahrenheit.
- `humid` –> relative humidity
- `wind_dir`, `wind_speed`, `wind_gust` –> wind direction (in degrees), speed and gust speed (in mph)
- `precip` –> precipitation in inches
- `pressure` –> sea level pressure in millibars
- `visib` –> visibility in miles
- `time_hour` –> date and hour (based on the year, month, day,hour fields)formatted as `YYYY-mm-dd HH:MM:SS`.

Data Cleansing

Due to a bug in the dataset, the data in the `time_hour` column are incorrectly shifted by 1 hour, so rectifying the dataset, using DAX expression that creates a new column.

time_hour_updated = weather[time_hour]-TIME(1,0,0)

Converting columns so that they can use metric (International System of Units, SI) or derived units.

- Converting columns `temp` and `dewp`, temperature and dew point in degrees Fahrenheit to degrees Celsius.
  - *temp_celsius = (weather[temp]-32) * 5/9*

- o *dewp_celsius = (weather[dewp]-32) * 5/9*

Name	dewp_celsius	Format	Decimal number	Summarization	Don't summarize
Data type	Decimal number	$ % , .00	2	Data category	Uncategorized
	Structure		Formatting		Properties

```
1 dewp_celsius = (weather[dewp]-32) * 5/9
```

- Converting column `precip`, precipitation in inches to millimetres.
  - o *precip_millimetres = weather[precip] * 25.4*

Name	precip_millimetres	Format	Decimal number	Summarization	Don't summarize
Data type	Decimal number	$ % , .00	2	Data category	Uncategorized
	Structure		Formatting		Properties

```
1 precip_millimetres = weather[precip] * 25.4
```

- Converting column `visib`, visibility in miles to metres.
  - o *visib_metres = weather[visib] * 1609.34*

Name	visib_metres	Format	Decimal number	Summarization	Don't summarize
Data type	Decimal number	$ % , .00	2	Data category	Uncategorized
	Structure		Formatting		Properties

```
1 visib_metres = weather[visib] * 1609.34
```

- Converting columns `wind_speed` and `wind_gust`, wind speed and wind gust speed in mph to mps.
  - o *wind_speed_mps = weather[wind_speed] * 0.44704*

Name	wind_speed_mps	Format	Decimal number	Summarization	Don't summarize
Data type	Decimal number	$ % , .00	2	Data category	Uncategorized
	Structure		Formatting		Properties

```
1 wind_speed_mps = weather[wind_speed] * 0.44704
```

  - o *wind_gust_mps = weather[wind_gust] * 0.44704*

Name	wind_gust_mps	Format	Decimal number	Summarization	Don't summarize
Data type	Decimal number	$ % , .00	2	Data category	Uncategorized
	Structure		Formatting		Properties

```
1 wind_gust_mps = weather[wind_gust] * 0.44704
```

- Creating column for `wind_speed_mps` with updating max outlier as 0.
  - o *wind_speed_mps_wo_max_outlier = IF(weather[wind_speed_mps] <> MAX(weather[wind_speed_mps]), weather[wind_speed_mps], 0)*

Name	wind_speed_mps_w...	Format	Decimal number	Summarization	Don't summarize	Sort by column	Data groups	Manage relationships
Data type	Decimal number	$ % , .00	2	Data category	Uncategorized			
	Structure		Formatting		Properties	Sort	Groups	Relationships

```
1 wind_speed_mps_wo_max_outlier = IF(weather[wind_speed_mps] <> MAX(weather[wind_speed_mps]), weather[wind_speed_mps], 0)
```

Data Visualization

- Now we will be using LGA airport from the 'origin' column and will be calculating daily mean of 'wind_speed'.

year	month	day	wind_speed
2013	1	1	6.69
2013	1	2	6.43
2013	1	3	4.91
2013	1	4	6.88
2013	1	5	5.14
2013	1	6	3.71
2013	1	7	4.76
2013	1	8	3.69
2013	1	9	2.57
2013	1	10	6.56
2013	1	11	3.17
2013	1	12	3.13
2013	1	13	2.53
2013	1	14	2.94
2013	1	15	3.88
2013	1	16	3.26
2013	1	17	4.59
2013	1	18	5.19
2013	1	19	5.79
2013	1	20	7.44

- As using above data values, will be plotting line plot, which is a graph that uses lines to connect data points to display data as points or check marks above a number line, showing the frequency of each value.

Average of wind_speed_mps by day
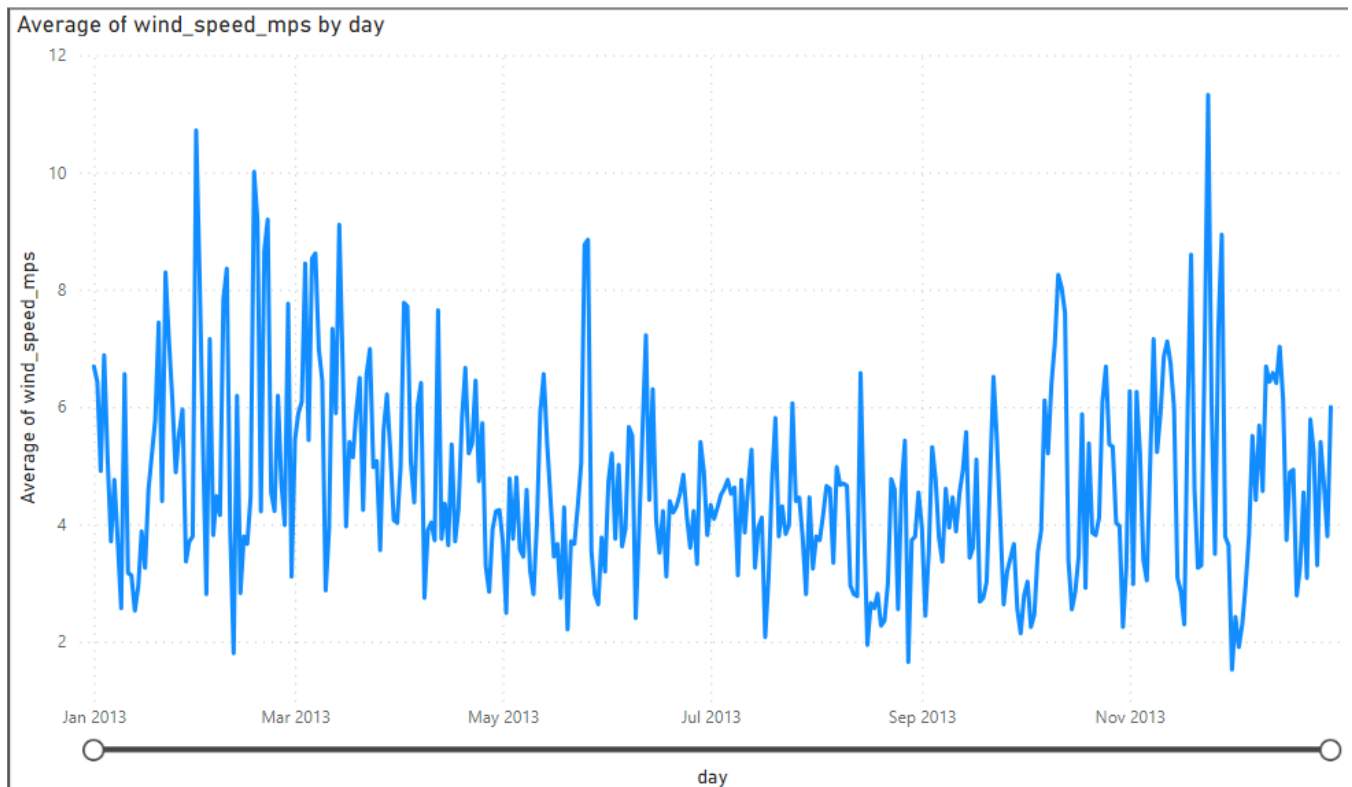
As per the visualization, we can derive:

- In 2013, daily average wind speed ranges from 1.5 to 11.5.
- Very high wind speed may lead to delay in flight.

- Fetching 10 windiest days at LGA airport with dates and the corresponding mean of daily wind speeds.

| date	Average of wind_speed_mps	Average of precip_millimetres
2013-11-24	11.32	0.00
2013-01-31	10.72	0.48
2013-02-17	10.01	0.00
2013-02-21	9.19	0.00
2013-02-18	9.17	0.00
2013-03-14	9.11	0.00
2013-11-28	8.94	0.00
2013-05-26	8.85	0.00
2013-05-25	8.77	0.08
2013-02-20	8.66	0.00

- Fetching monthly mean of wind speed for all the 3 airports.

origin	month	year	Average of wind_speed_mps	month_year
EWR	1	2013	4.33	2013-01
JFK	1	2013	5.38	2013-01
LGA	1	2013	5.07	2013-01
EWR	2	2013	4.72	2013-02
JFK	2	2013	5.95	2013-02
LGA	2	2013	5.52	2013-02
EWR	3	2013	5.14	2013-03
JFK	3	2013	6.21	2013-03
LGA	3	2013	5.85	2013-03
EWR	4	2013	4.27	2013-04
JFK	4	2013	5.58	2013-04
LGA	4	2013	4.93	2013-04
EWR	5	2013	3.71	2013-05
JFK	5	2013	4.60	2013-05
LGA	5	2013	4.20	2013-05
EWR	6	2013	4.20	2013-06
JFK	6	2013	4.88	2013-06
LGA	6	2013	4.46	2013-06

- As we have created a new dataset of monthly mean of wind speeds for all the 3 airports, now we will be plotting a graph using line plot, which is a graph that uses lines to connect data points to display data as points or check marks above a number line, showing the frequency of each value.

monthly average wind speed [m/s] by month and origin
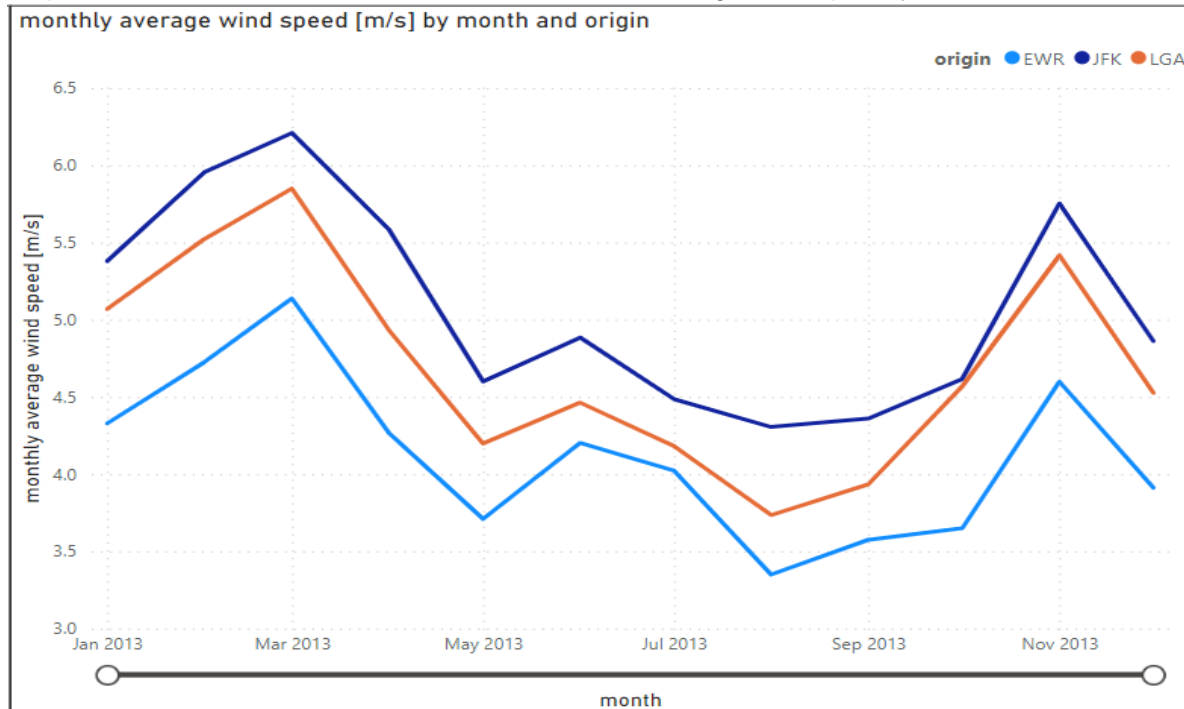
As per the visualization, we can derive:

- `JFK` origin flights have encountered more average wind speed than other airports.
- `EWR` origin flights have encountered less average wind speed than other airports.

Dashboard Link

G-drive Link: https://drive.google.com/file/d/1xstCbVtuNMeGPO_bd7NHEalMkHAIJeTJ/view

Power BI Link: https://app.powerbi.com/groups/me/reports/6815e94a-51d1-4c27-8449-5cb60794e5a3/ReportSection?experience=power-bi

Summary

In this report we have worked with the flights departing from NYC in 2013. Over this dataset we have done some exploratory data analysis, before which have pre-processed the data by modifying units and rectifying time. Using the final dataset, calculated the mean and used it with different visualization for wind checks for different airports.

In our future analysis we can add up more features, via which we can know timings, delay, reason for delay and more and how these features are affecting the other features for further predictions.

References

No references have been used for this report, except video references from courses.