



DATA DRIVEN

AIRLINE COMPLAINTS ANALYSIS

ABISHEK K S
SRI SAIRAM ENGINEERING COLLEGE
08-10-2023

INTRODUCTION

Air carriers report flight delays, cancellations, overbookings, late arrivals, baggage complaints, and other operating statistics to the U.S. government, which compiles the data and reports it to the public.

The data set contains monthly observations from 2004 to 2010 for United Airlines, American Eagle, and Hawaiian Airlines. The variables in the data set include:

- **Baggage** - The total number of passenger complaints for theft of baggage contents, or for lost, damaged, or misrouted luggage for the airline that month
 - **Scheduled** - The total number of flights scheduled by that airline that month
 - **Cancelled** - The total number of flights cancelled by that airline that month
 - **Enplaned** - The total number of passengers who boarded a plane with the airline that month.
- A small python script was executed to get the following information about the dataset. It is observed that the dataset is clean with no NULL objects with proper formatting. Hence, we proceed with our analysis steps.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 252 entries, 0 to 251
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Airline      252 non-null    object
1   Date         252 non-null    object
2   Month        252 non-null    int64
3   Year         252 non-null    int64
4   Baggage      252 non-null    int64
5   Scheduled    252 non-null    int64
6   Cancelled    252 non-null    int64
7   Enplaned     252 non-null    int64
dtypes: int64(6), object(2)
memory usage: 15.9+ KB
None
```

	Month	Year	Baggage	Scheduled	Cancelled	Enplaned
count	252.000000	252.000000	252.000000	252.000000	252.000000	2.520000e+02
mean	6.500000	2007.000000	12613.507937	28128.007937	703.761905	2.203871e+06
std	3.458922	2.00398	9993.307166	17092.087874	746.020368	1.788200e+06
min	1.000000	2004.000000	1033.000000	3553.000000	0.000000	4.234460e+05
25%	3.750000	2005.000000	1910.500000	5565.750000	25.750000	6.865205e+05
50%	6.500000	2007.000000	12224.000000	36696.000000	533.000000	1.391112e+06
75%	9.250000	2009.000000	19359.250000	42162.500000	1078.500000	4.111049e+06
max	12.000000	2010.000000	41787.000000	50837.000000	3712.000000	6.137271e+06

```
Airline      Date  Month  Year  Baggage  Scheduled  Cancelled  Enplaned
```

```
import matplotlib.pyplot as plt

import ydata_profiling as pp
from ydata_profiling import profile_report
import pandas as pd

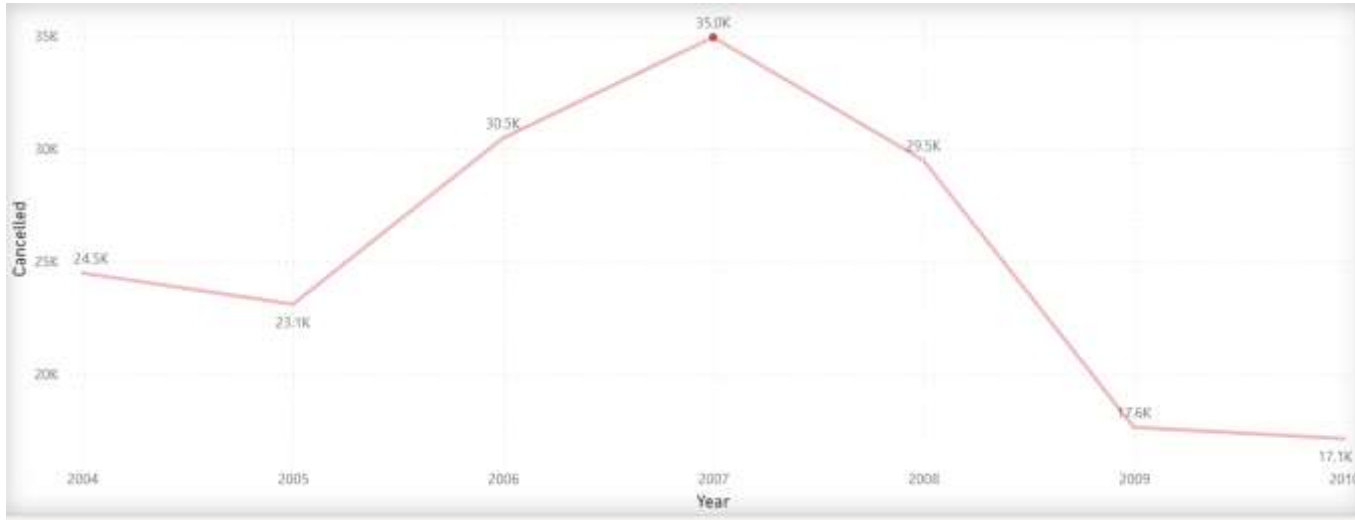
data = pd.read_csv("baggagecomplaints.csv")

ppd=pp.ProfileReport(data)
ppd.to_file('index.html')
print(data.head())
print('-----')
print(data.info())
print('-----')
print(data.describe())
print('-----')
print(data.tail())
```

OBJECTIVES

1. To gain insights on the four metrics [**Baggage, Scheduled, Cancelled, Enplaned**] for all the three given airlines [**American Eagle, Hawaiian, United**]
2. To rank the given airlines according to the lowest number of complaints received.
3. To gain insights on the trends of baggage-related complaints and elucidate their trends over the given years.
4. To find out the 'Best Months' for every given airline.

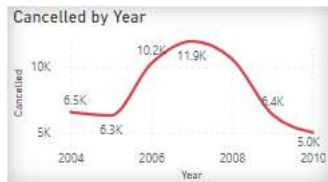
Trends of Cancellations



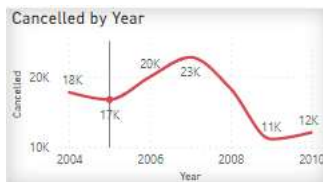
The year 2007 has seen the most number of cancellations in all the given years.

1. From a glance, the values seem to take the form of a normal distribution. Two tests were performed on the given data range.
2. Since visual tests look similar to a Normal Distribution, Shapiro Wilk's test was conducted on the given array.
3. The computed p-value stood at 0.64, which was far less than 0.05 and hence the Cancellation trends are normally distributed across all airlines.

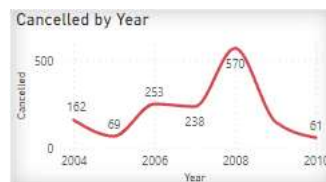
```
[2004, 2005, 2006, 2007, 2008, 2009, 2010] [24497, 23111, 30481, 34959, 29507, 17647, 17146]  
ShapiroResult(statistic=0.9411253929138184, pvalue=0.6488763689994812)  
Data is normally distributed.
```



United Airlines



American Eagle



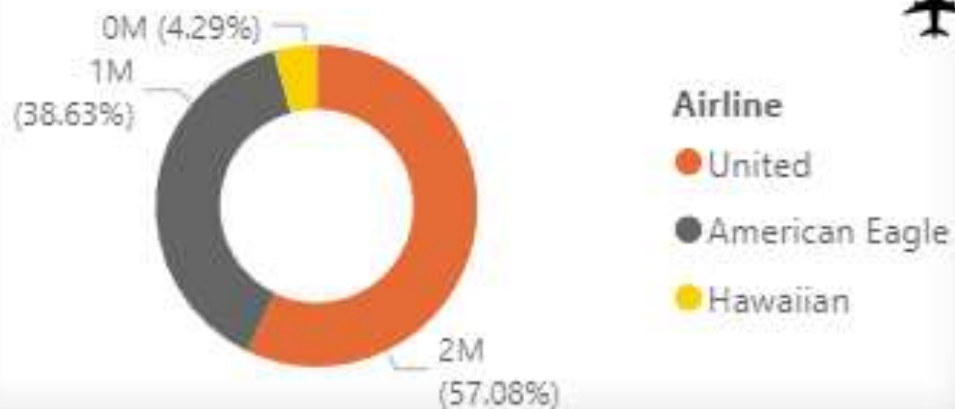
Hawaiian Airlines

Normal Distribution in all cases

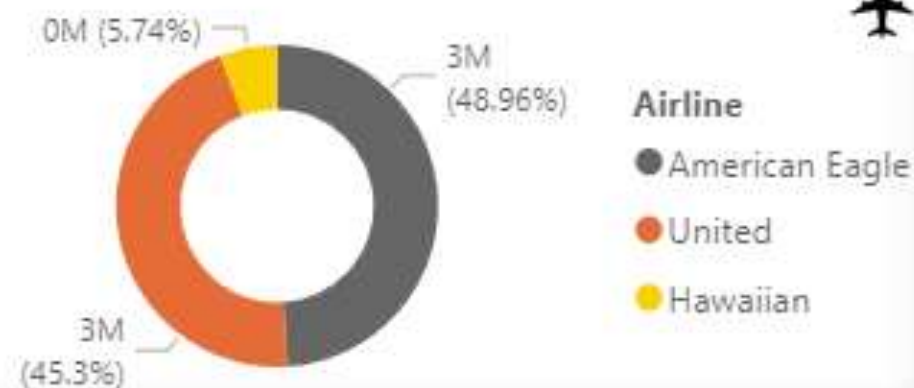
```
Data is normally distributed.  
Am-Eag Data is normally Distributed  
Hawaii Data is normally Distributed  
United Data is normally Distributed
```


PARAMETRIC SHARES.

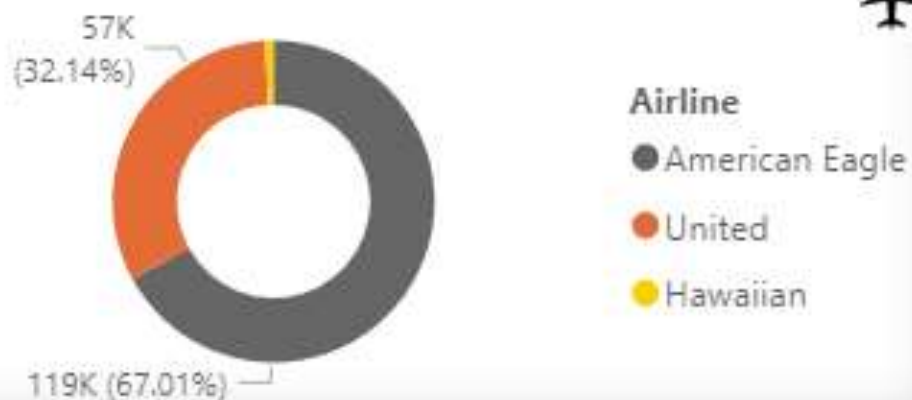
Total Baggage by Airline



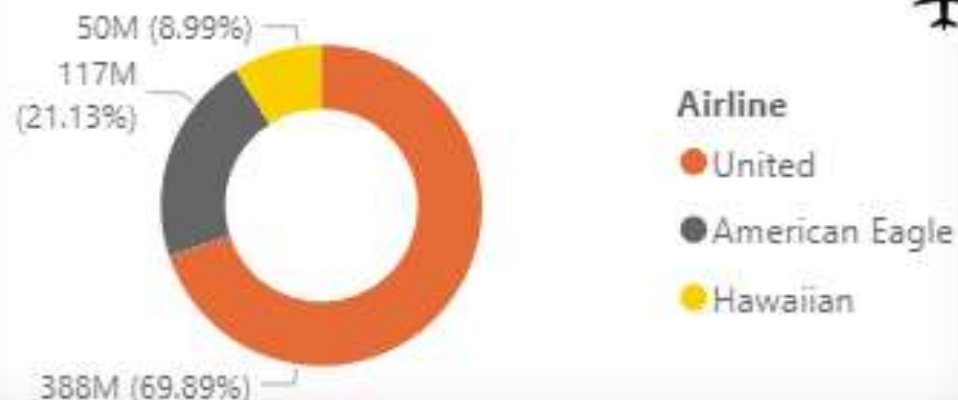
Total Scheduled by Airline



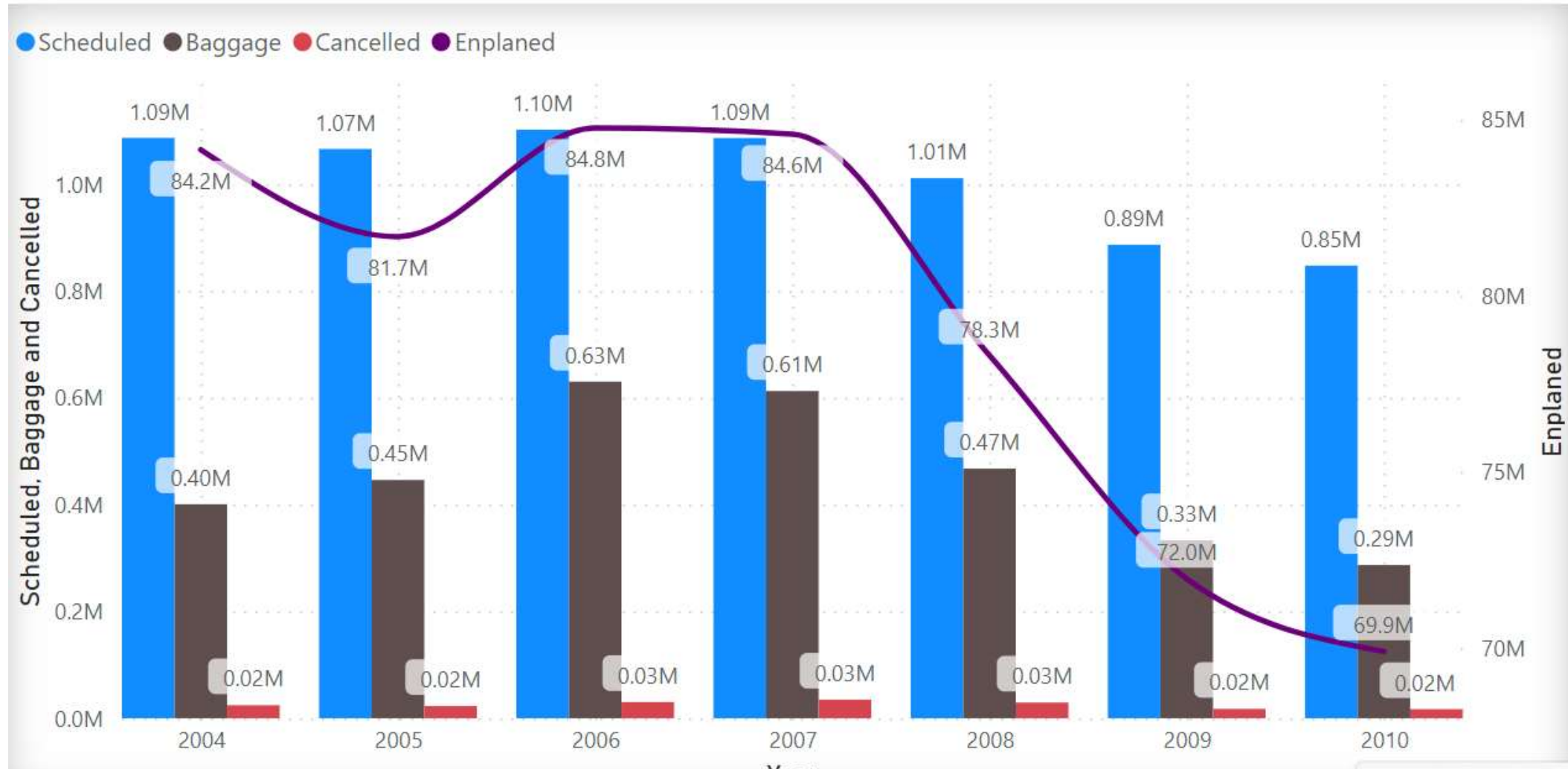
Total Cancelled by Airline



Total Enplanned by Airline



DATA DRIVEN



The four parameters can be summarized in a single visual as shown.

The trends in baggage complaints clearly follow Gaussian Distribution.

- **United Airlines** has seen the highest number of baggage related complaints over the years as compared to other airlines.
- **United Airlines** has seen 1,814,380 [Close to 2 Million] Baggage related complaints, comprising of 57.08% of the total baggage related complaints, followed by **1,228,103 complaints on American Eagle Airlines [38.63% of total]** and **Hawaiian Airlines** with the **least number** of complaints at 136,211 [4.30% of total].
- **American Eagle** has cancelled the most number of airlines across all the years, followed by United Airlines and Hawaiian Airlines.
- United - 388,139,830, American Eagle - 117,324,946 and Hawaiian - 49,910,630 are the Enplanned numbers across the years.

But since the above analysis is just based on total numbers, a more capacity-centered metric would be appropriate.

Since Enplanned ratios are different, it is better to compute a Scheduled: Cancelled ratio to assess performance.

- I. American Eagle has a S:C value of 0.00685, which means approximately **7 flights** out of 1000 flights is cancelled.
- II. Hawaiian Airlines has the S:C Value at **0.0030113**, which means **3 flights** are cancelled out of every 1000 flights.
- III. United Airlines has an S:C value at **0.014**, which means **14-15 flights** are cancelled out of every 1000 flights.

Thus, United Airlines Cancels more flights than the other two airlines combined.

DATA DRIVEN

December and January months have seen the greatest number of cancellations, averaging around 1000 Cancellations across all the years.

The general trend of cancellations is that of a downward trend from Feb→November.

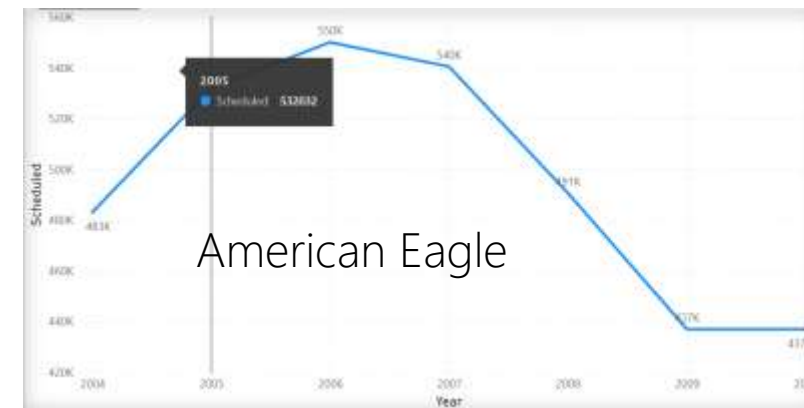
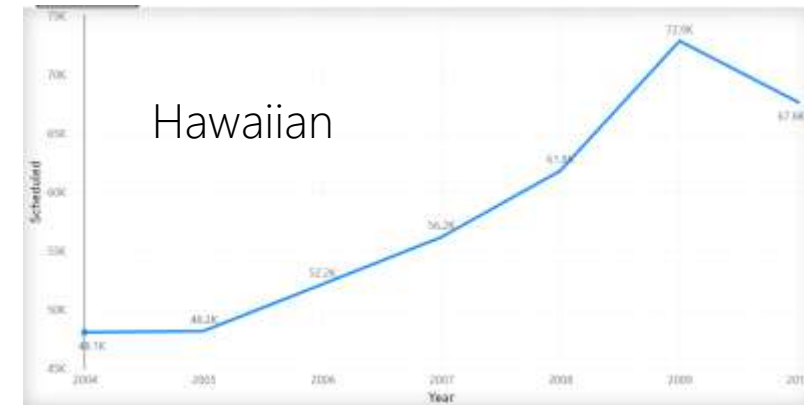
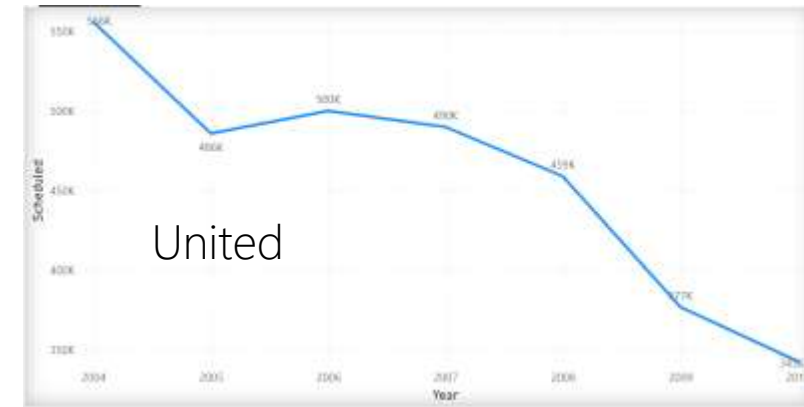
Hawaiian airlines doesn't follow this general trend. The airline has seen more cancellations in the month of December.

Most number of Cancellations:

- Hawaiian airlines has seen the greatest number of cancellations in 2008.
- United airlines has seen the greatest number of cancellations in 2007.
- American Eagle has seen the greatest number of cancellations in 2007

SCHEDULING TRENDS

- I. United Airlines has seen a downward trend, that is, United Airlines has reduced the number of scheduled flights. This might be an indicator of downward performance.
- II. Hawaiian airlines has seen an upward trend in scheduling.
- III. American Eagle Scheduling trends have followed a nearly-perfect gaussian distribution, but it's scheduled flight count has remained nearly same as that of 2006.
- IV. January and December months have seen the highest number of schedulings for both United Airlines and American Eagle.
- V. July, August and December months have seen the greatest number of schedulings for Hawaiian airlines.



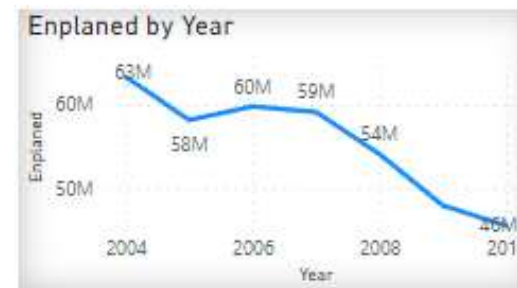
ENPLANNED TRENDS

SUMMARY:

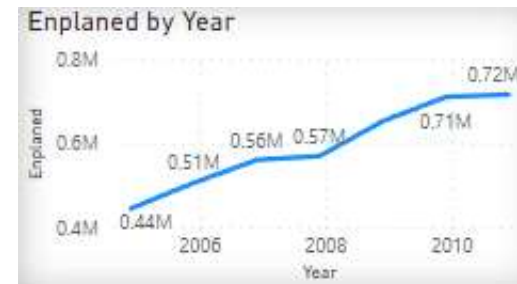
The number of people who choose to travel with United Airlines has decreased over the years.

Hawaiian Airlines has seen an increase in the number of people over the years.

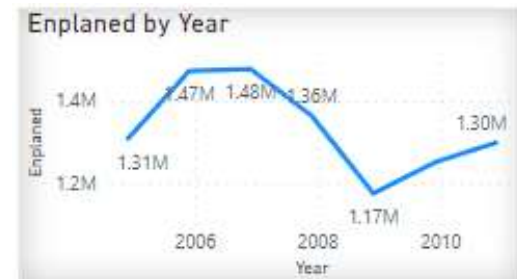
The performance of American Eagle has remained somewhere in between.



United Airlines



Hawaiian Airlines



American
Eagle

CONCLUSION

- I. Hawaiian Airlines has been performing significantly better and is improving over the years.
- II. United Airlines has seen a downward trend in performance with more cancellations and reduced number of people travelling with it.
- III. American Eagle has been improving over the years.

REF :

Dataset: <https://www.kaggle.com/datasets/gabrielsantello/airline-baggage-complaints-time-series-dataset>

Dashboard: <https://www.novypro.com/project/airline-complaints-dashboard>