

== UJIAN AKHIR SEMESTER ==

= LAPORAN ANALISIS DATA =

NPM>Nama Mahasiswa/Kelas: 062140832957 Muhammad Alif Fauzan Hanafi 5-MIM

Nama Dataset: AirlineDelayCause

Informasi Yang Akan Digali Dari Dataset:

Dari dataset ini menggali informasi delay dari alasan cuaca dan keamanan serta jumlah delay yang terjadi bahkan diluar alasan tadi.

Langkah Proses Preprocessing Data:

- **Jumlah sample data Finansial tiap klasifikasi**

Drag data security, weather, dan airline delay cause dari tables ke columns lalu airport serta carrier name ke rows.

```
[2] import pandas as pd
df = pd.read_csv('Employee Sample Data.csv', encoding='windows-1252')
df.shape

(1000, 14)

[3] data_na = (df.isnull().sum() / len(df)) * 100
data_na = data_na.drop(data_na[data_na == 0].index).sort_values(ascending=False)[:30]
missing_data = pd.DataFrame({'Missing Ratio' :data_na})
missing_data.head(20)

Missing Ratio
Exit Date      91.5

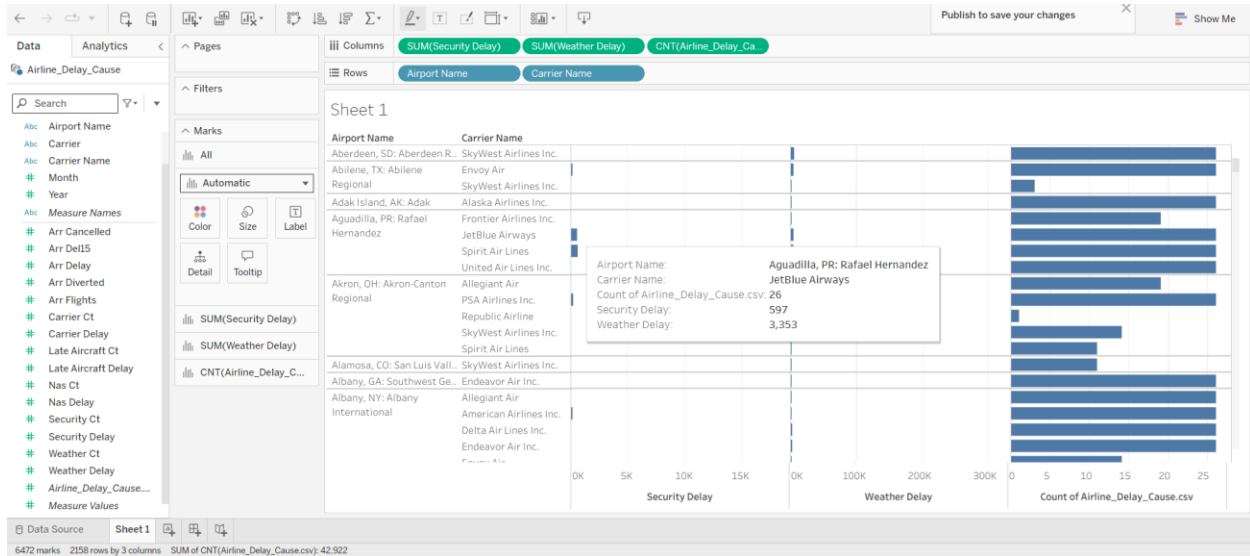
[4] duplicate_rows_data = df[df.duplicated()]
print("number of duplicate rows: ", duplicate_rows_data.shape)

number of duplicate rows: (0, 14)

[5] df.drop(0, axis=0, inplace=True)
```

Hasil Visualisasi:

- Jumlah delay dan penyebabnya



Kesimpulan Dari Visualisasi:

Dapat disimpulkan dari visualisasi data yang dilakukan maka didapatkan deskripsi informasi detail tentang jumlah delay yang terjadi juga yang spesifik dari alasan cuaca dan keamanan.