

# **== UJIAN AKHIR SEMESTER ==**

## **= LAPORAN ANALISIS DATA =**

***NPM>Nama Mahasiswa/Kelas: 062140830489 Muhammad Husaini Hasyim 5-MIM***

***Nama Dataset: EmployeeSampleData***

### ***Informasi Yang Akan Digali Dari Dataset:***

Dari dataset ini didapatkan klasifikasi informasi yang saya buat yaitu rincian Umur dan Gaji Pegawai, Posisi dengan Bonusnya, dan Negara beserta jumlah Pegawainya.

### ***Langkah Proses Preprocessing Data:***

- **Umur dan Gaji**

Drag data **Full Name** dari tables ke columns dan drag data **Gender, Age**, lalu **Annual Salary** ke rows

- **Posisi dan Bonus**

Drag data **Bonus** dari tables ke columns dan drag data **Department**, lalu **Job Title** ke rows

- **Negara beserta jumlah Pegawai**

Tarik data **Country** dari tables ke field sheet dan drag data **Count of Employee Sample Data** ke Marks

01>Loading\_and\_preprocessing\_your\_own\_data.ipynb

File Edit Lihat Sisipkan Runtime Fitur Bantuan Tidak dapat menyimpan perubahan

+ Kode + Teks Salin ke Drive

### Loading and preprocessing your own data

In this tutorial, we demonstrate how to process your own dataset for later analysis on Google Colab.

We review the following steps:

1. Loading your own data
2. Filtering
3. Normalization
4. Transformation

```
[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

Employee Sample Data.csv

1 to 10 of 1000 entries

EID	Full Name	Job Title
E02387	Emily Davis	Sr. Manager
E04105	Theodore Dinh	Technical Architect
E02572	Luna Sanders	Director
E02832	Penelope Jordan	Computer Systems Manager
E01639	Austin Vo	Sr. Analyst
E00644	Joshua Gupta	Account Representative
E01550	Ruby Barnes	Manager
E04332	Luke Martin	Analyst
E04533	Easton Bailey	Manager
E03838	Madeline Walker	Sr. Analyst

Show 10 per page

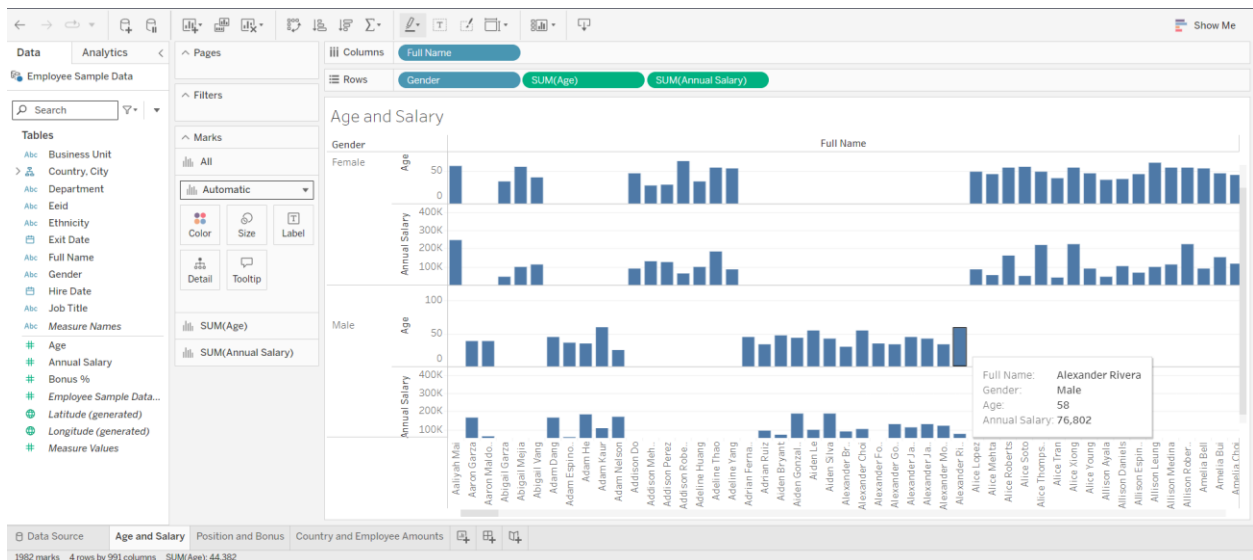
```
[4] duplicate_rows_data = df[df.duplicated()]
print("number of duplicate rows: ", duplicate_rows_data.shape)

number of duplicate rows: (0, 14)
```

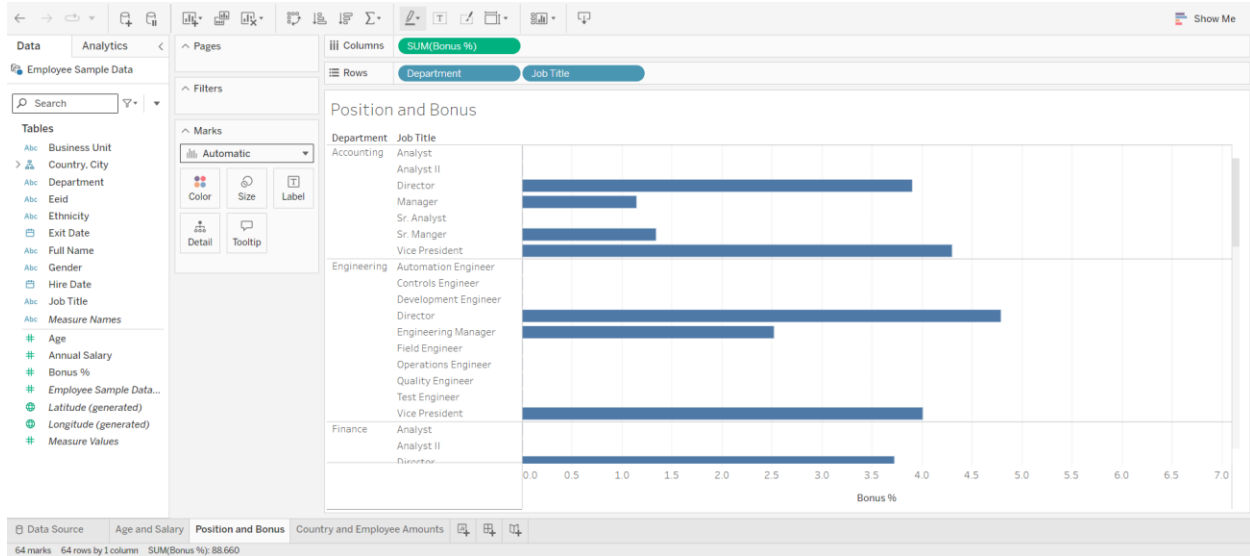
```
[5] df.drop(0, axis=0, inplace=True)
```

*Hasil Visualisasi:*

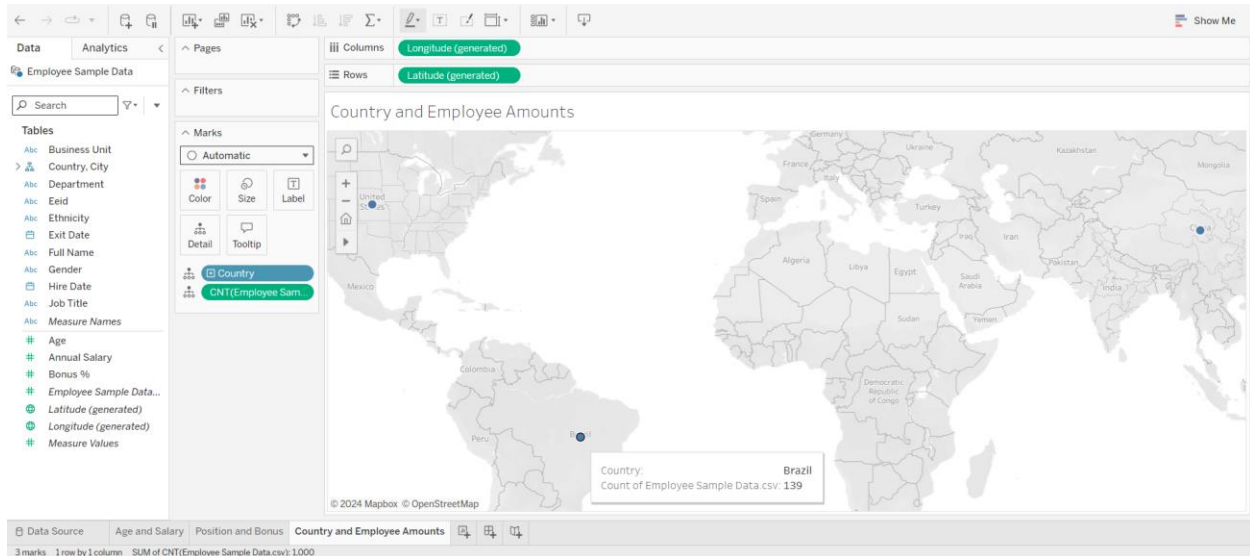
## - Umur dan Gaji Pegawai



## - Posisi dan Bonus



## - Negara beserta jumlah Pegawai



### *Kesimpulan Dari Visualisasi:*

Dapat disimpulkan dari visualisasi data yang dilakukan maka didapatkan deskripsi informasi detail terkait dengan klasifikasinya terbagi jadi detail data umur dengan gaji pegawai, lalu posisi dengan bonusnya, dan terakhir masing-masing negara dengan jumlah pegawai. Dari dataset ini di tiap sheet.