# Exploratory Data Analysis (EDA) for Mercedes-Benz San Francisco on the SF Salaries dataset.

## Abstract:

This project conducted for T5 data science boot camp, where we performed Exploratory Data Analysis (EDA) on the SF Salaries data. This project aims is to ensure your Mercedes-Benz car ownership experience exceeds your expectations, to develop marketing plans for our next marketing campaign, we decided to analysis a dataset of San Francisco employees' salary base to know income level and target each category with the appropriate promotional campaign.

## Design (Background Company):

- **Company info**: Mercedes-Benz San Francisco has been in the automotive industry for more than 50 years. Mercedes-Benz has been recognized as the benchmark in terms of high quality, technical models, luxury and value.

- **Problem/opportunity:** There is not enough information about the salaries of San Francisco employees to analyze it and make advertising and marketing decisions for Mercedes cars

- **Value for the company:** Based on the 10 most popular jobs among San Francisco residents, we will offer promotions to their incumbents and Depending on the salary level of the population of high, medium and low levels, we can target them with offers that suit their level, we also target the most widespread category, the middle-income category, by establishing facilities that make them own Mercedes cars.

## Data:

- **About SF salaries**: One way to understand how a city government works is by looking at employees and how it's employees compensated. This data contains the names, job title and pays compensation for San Francisco city.
- **Data description**:
  The original source for this data is here , and we have taken from kaggel (SF Salaries | Kaggle).
  This data set is name as Salaries contain 13 column and 148654 rows.

- **Column description**:
  **Id:** A unique identifier for each employee.
  **EmployeeName:** Name of employee.
  **JobTitle:** Names of Job.
  **BasePay:** Base salary without any additions.
  **OvertimePay:** Amount of overtime pay.
  **OtherPay:** Amount of other pay.
  **Benefits:** Amount of benefits.
  **TotalPay:** Amount of base pay plus overtime pay without benefits.
  **TotalPayBenefits:** Amount of base pay plus overtime pay plus benefits.
  **Year:** year for every jobs
  **Notes:** comment for job.
  **Agency:** city of job.
  **Status:** type of job (PT, FT).

## Algorithms:

1. Problem understanding.
2. Data validation (cleaning data):
   - Null value.
   - Outlier.
   - Upper to lower.
   - Negative value.
   - Duplicated rows.
3. Exploratory Data Analysis.
4. Questions & Insights.
5. Conclusion.

## Tolls:

- **Technology:**
  - Python.
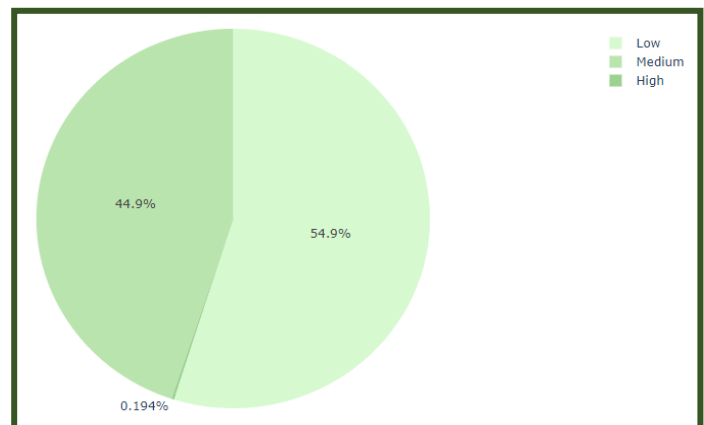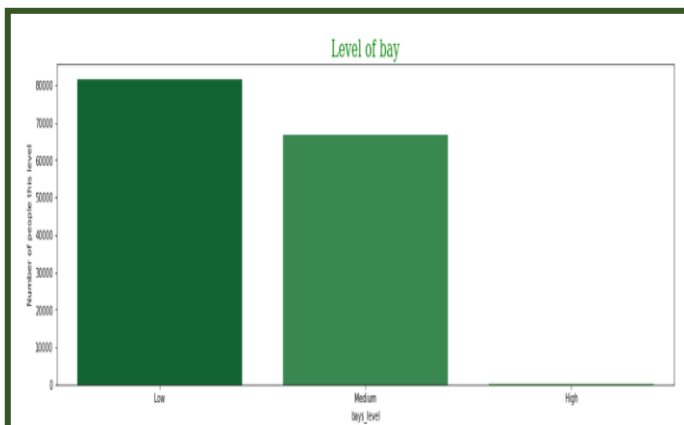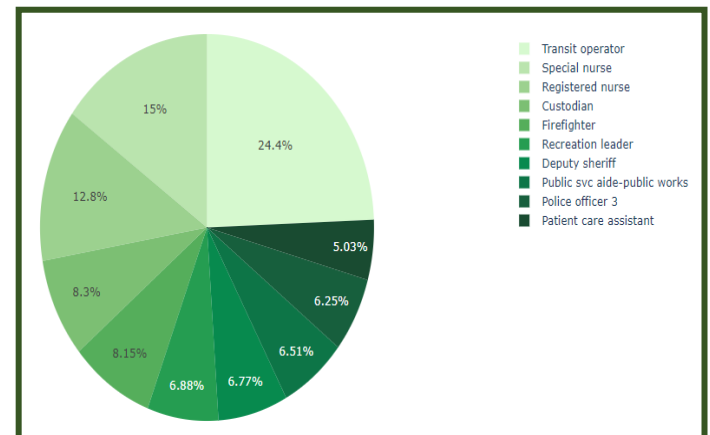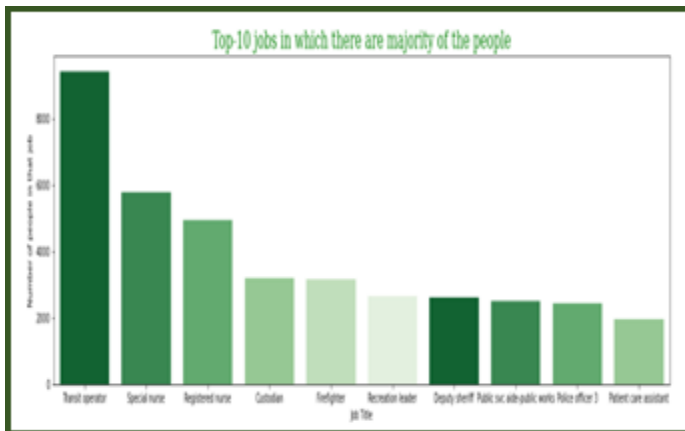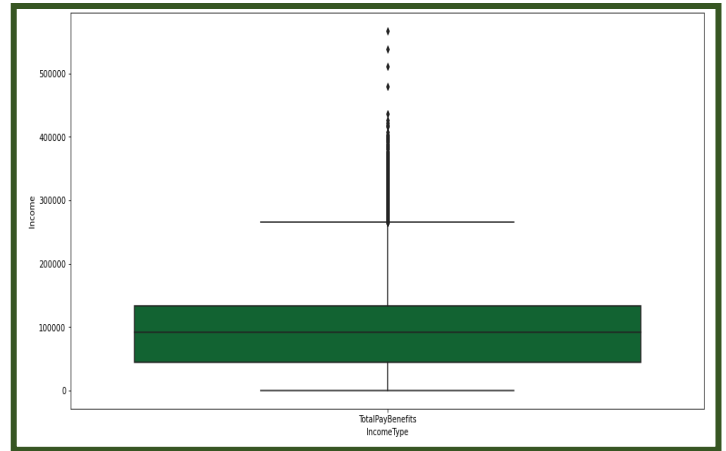  - Jupyter Note book.
- **Libraries:**
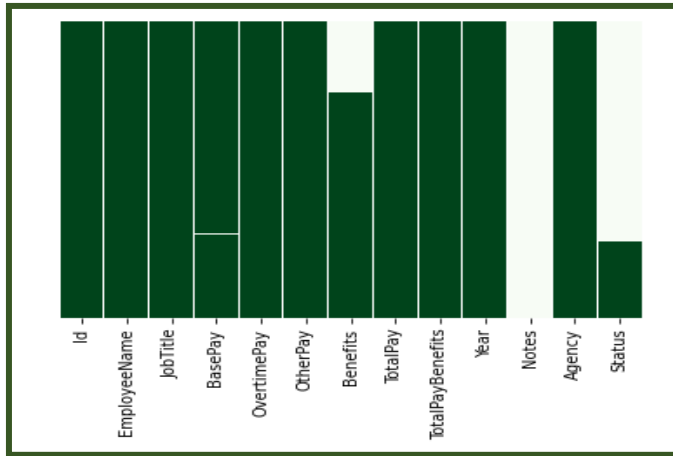  - NumPy and Pandas for data manipulation.
  - Matplotlib and Seaborn for plotting.
  - Plotly for interactive visualizations.

# Communication:

## -Charts:





### Top-10 jobs in which there are majority of the people





- Transit operator
- Special nurse
- Registered nurse
- Custodian
- Firefighter
- Recreation leader
- Deputy sheriff
- Public svc aide-public works
- Police officer 3
- Patient care assistant

### Level of bay





- Low
- Medium
- High

**-Presentation snips:**

[Press here to go into Presentation](#)