



SALARIES DATASET ANALYSIS REPORT

BY

A l i A m r A l i A b d a l l a

Basic Data Exploration



148654

Rows



13

Column



0

Duplicated

- There is 4 column with Null ["BasePay","OvertimePay","OtherPay","Benefits"] so we will fixed it .
- Notes and Status columns has no values in it so we will drop them.
- All Data types seems all right.

Missing Values

- **Missing values has taken a place in the data as there is 25% missing values in Benefits column.**
- **So we need to impute it rather than just drop it.**
- **I used the Mean values for imputing as it is the most common way to solve such a problem.**

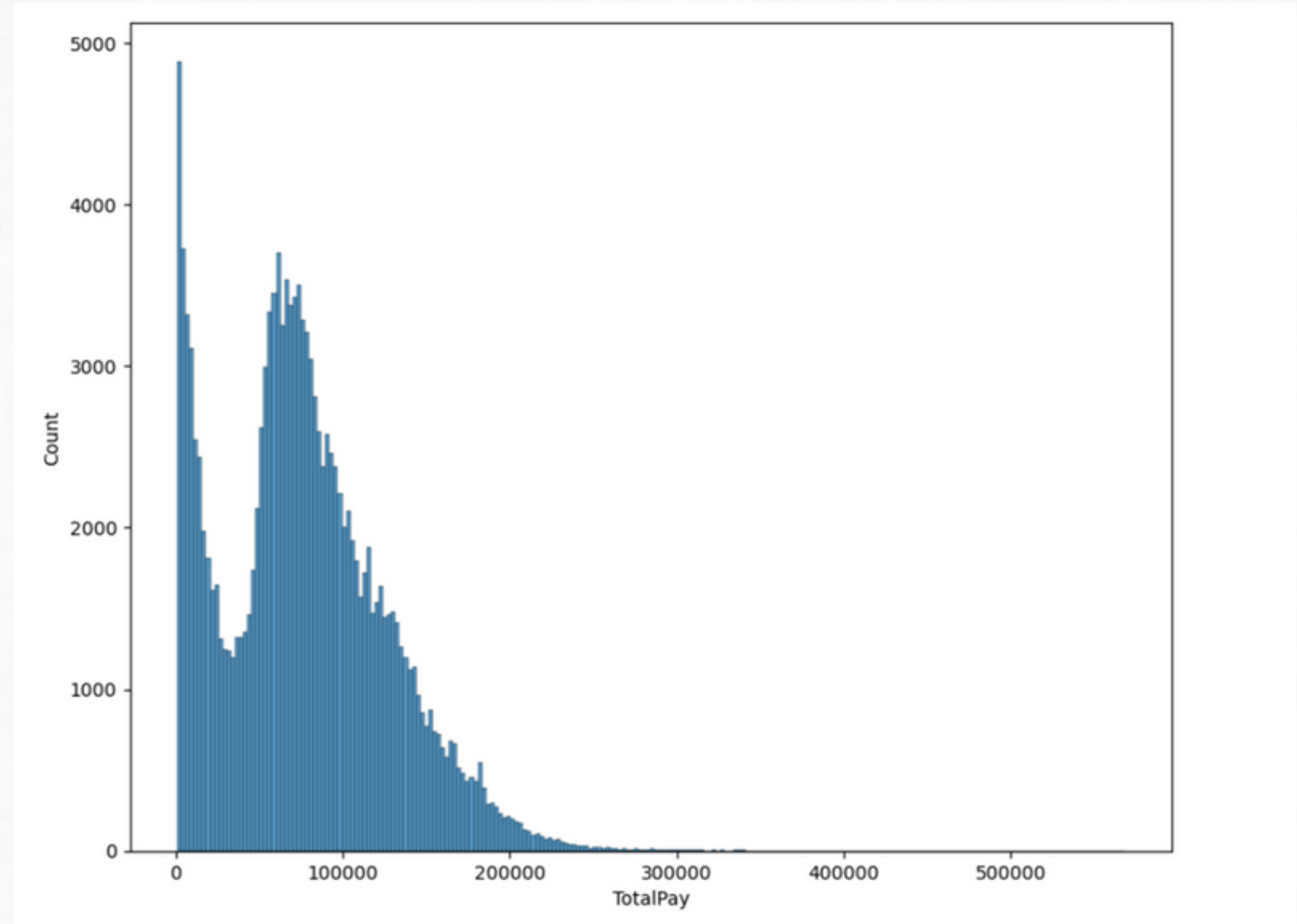
Descriptive Statistics

- **The Descriptive Statistics shows that, There is a -ve values in all the Payment column, which need to fix.**
- **There is rows with 0 values in BasePay, TotalPayBenefits, TotalPay columns, Which need to fix.**

Basic Data Visualization

Distribution of Salaries:

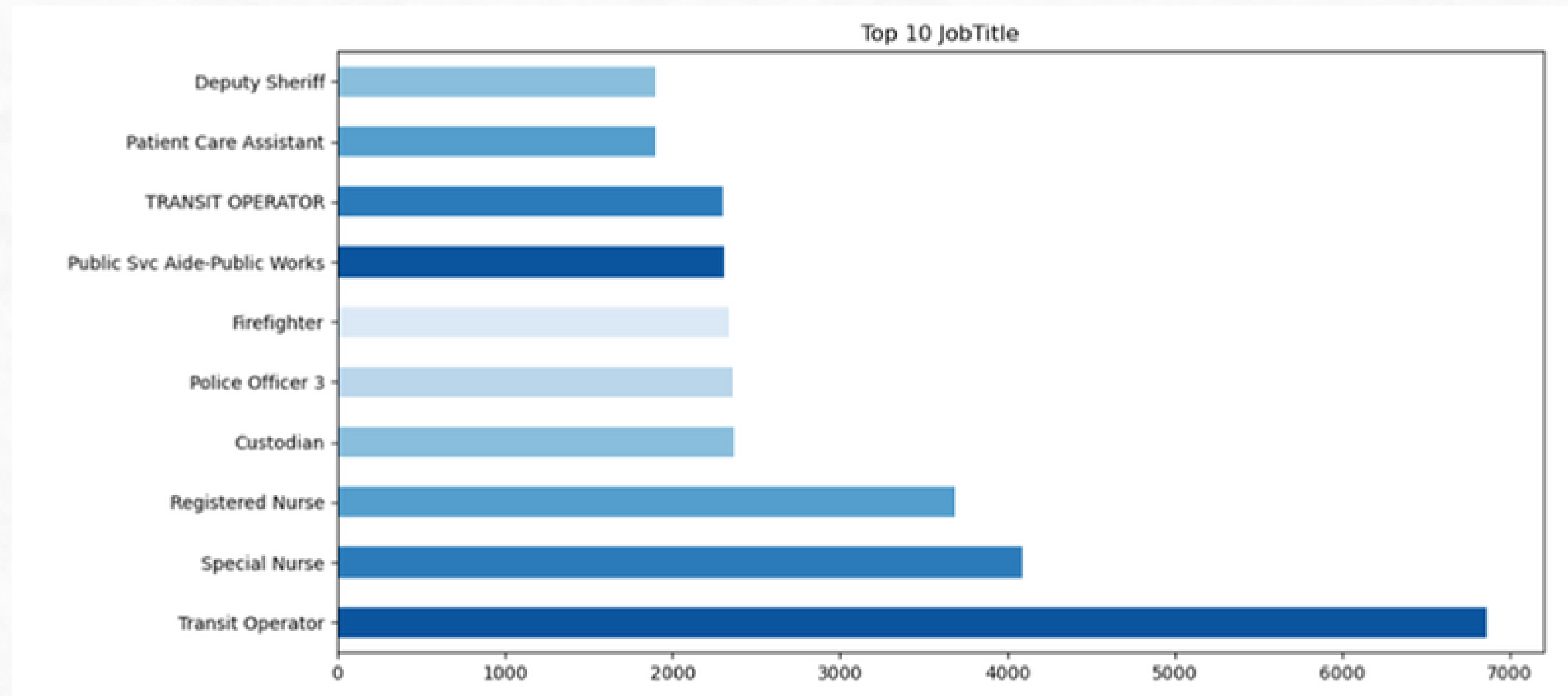
The Salary is'n a normal distribution, It's a Right Skewed



Basic Data Visualization

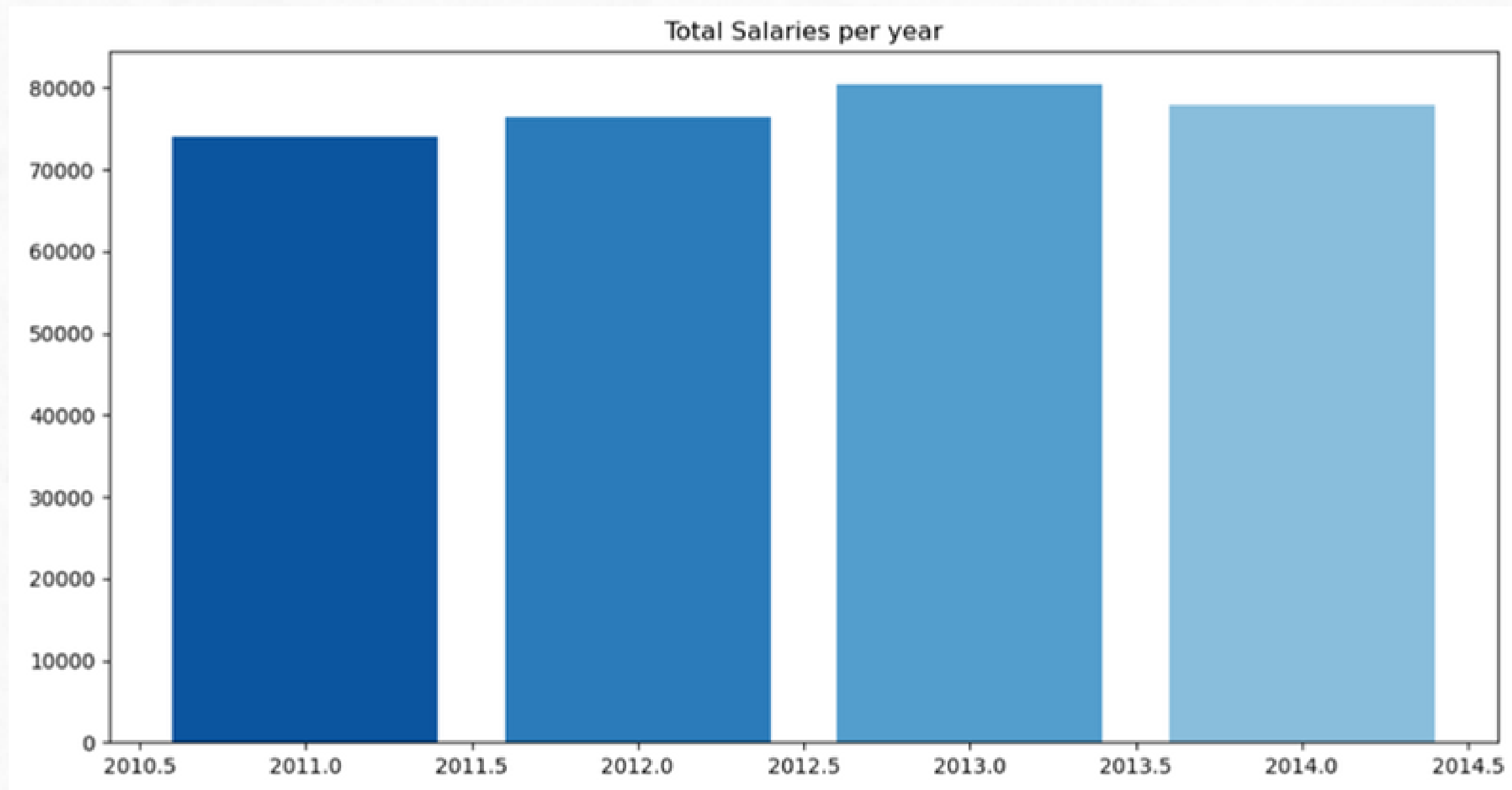
**Top 10 Jop_titles
platform:**

**Transit Operator is the
most frequnty job.**



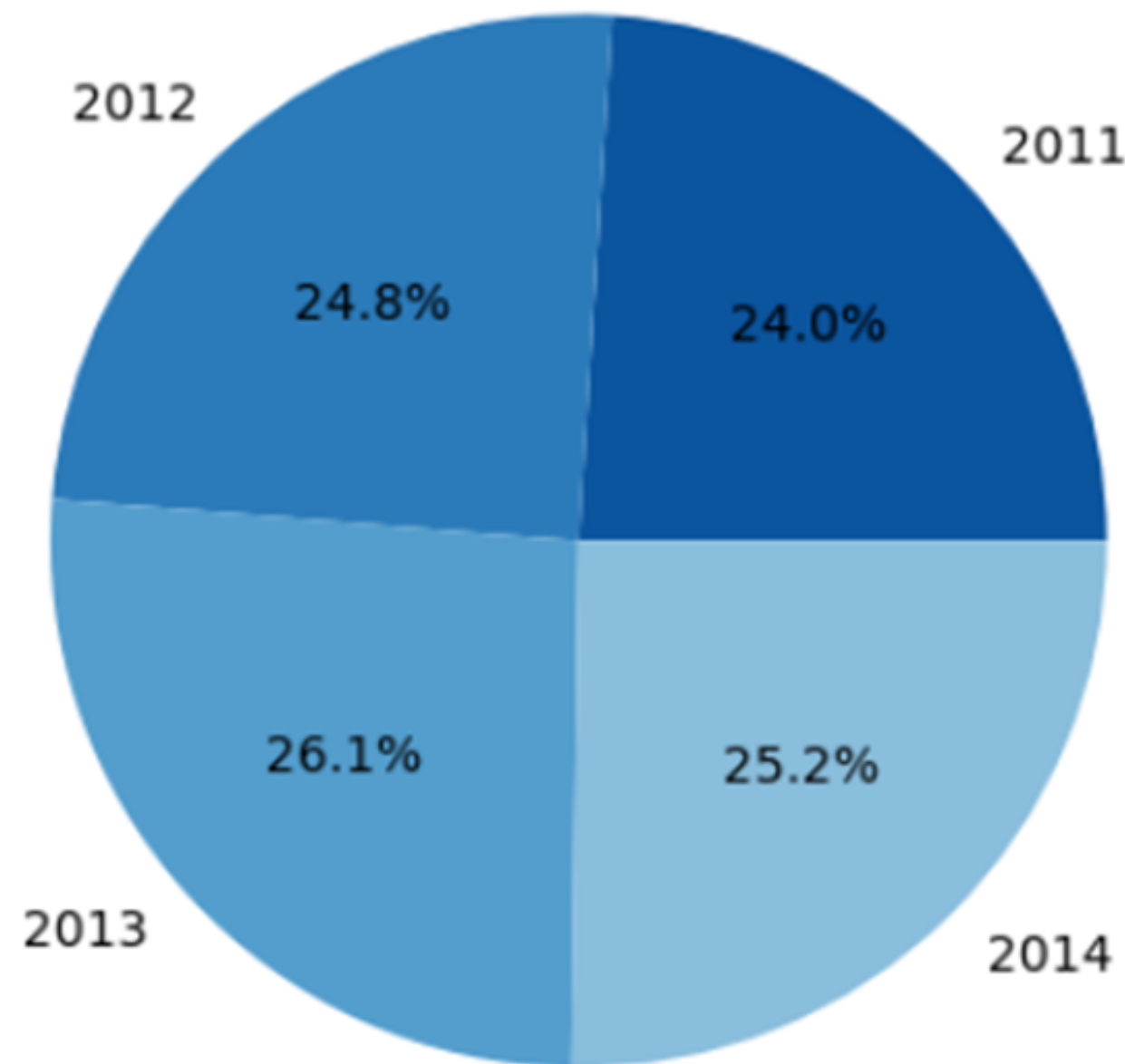
Basic Data Visualization

Seems like 2013 is the golden year for the employees



Basic Data Visualization

Seems like 2013 is the golden year for the employees

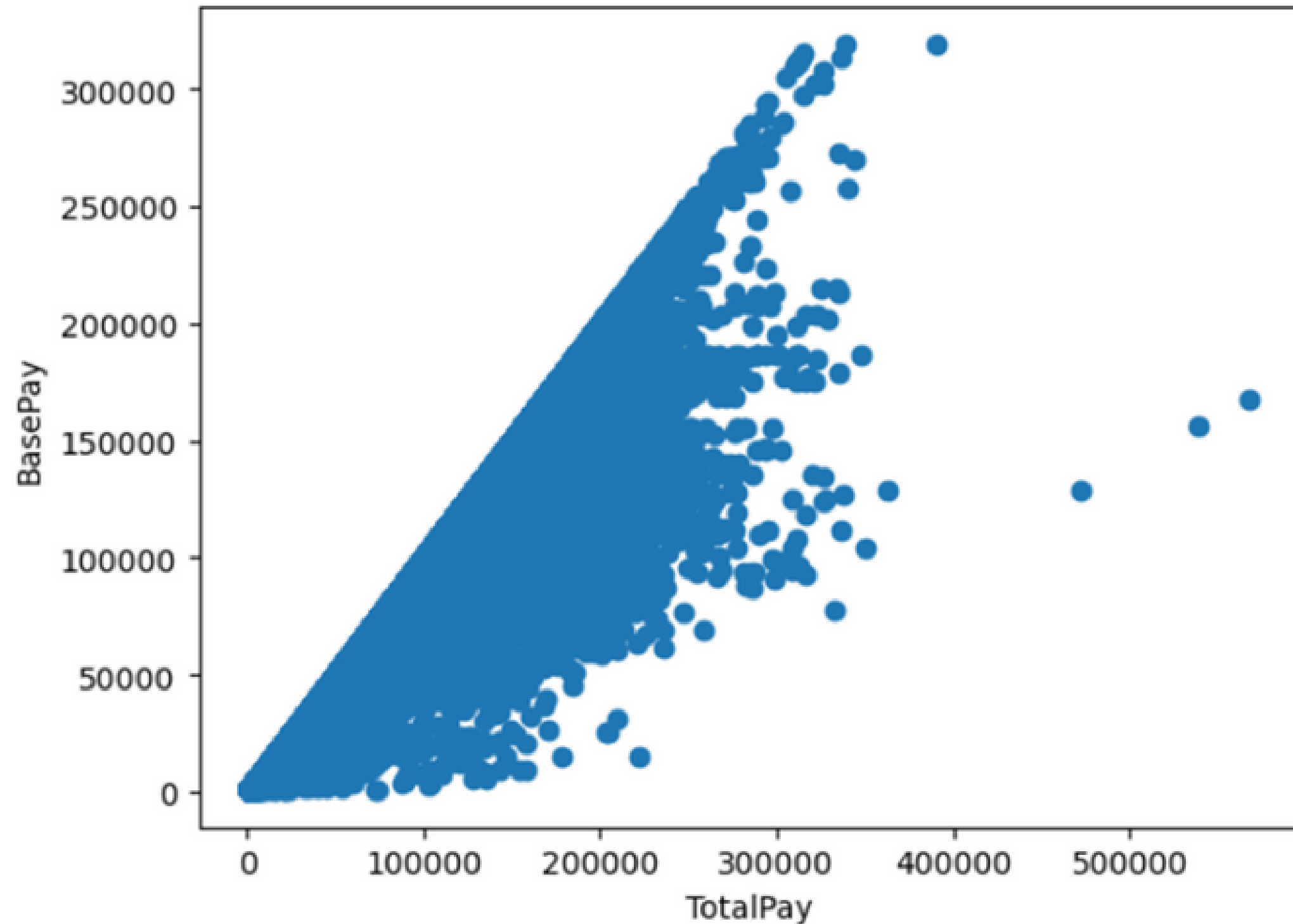


Simple Correlation Analysis

- **So as we calculate the Correlation for the TotalPay column.**
- **We Found that good correlation is with BasePay, Benefits and TotalPayBenefits**

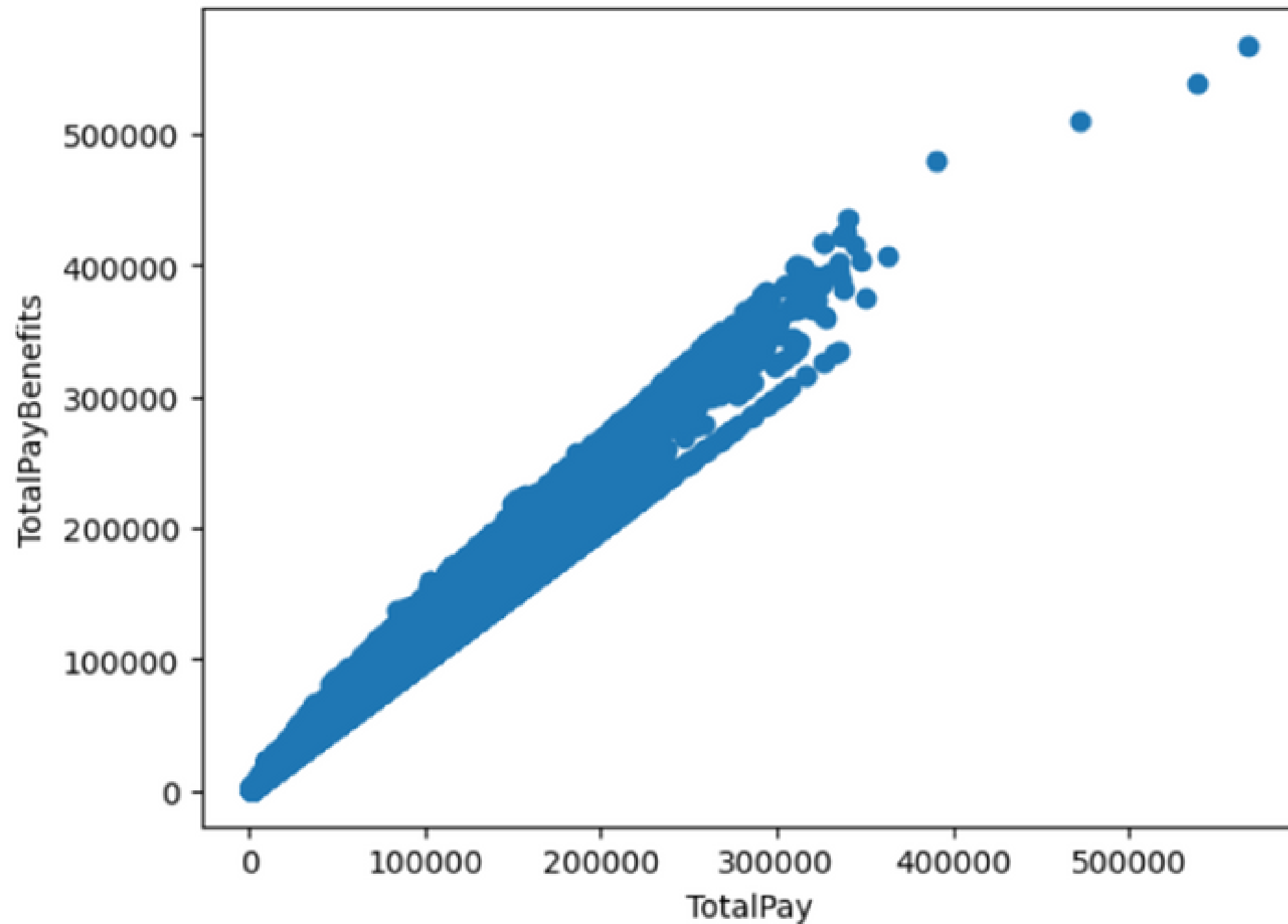
Simple Correlation Analysis

Correlation between TotalPay and BasePay



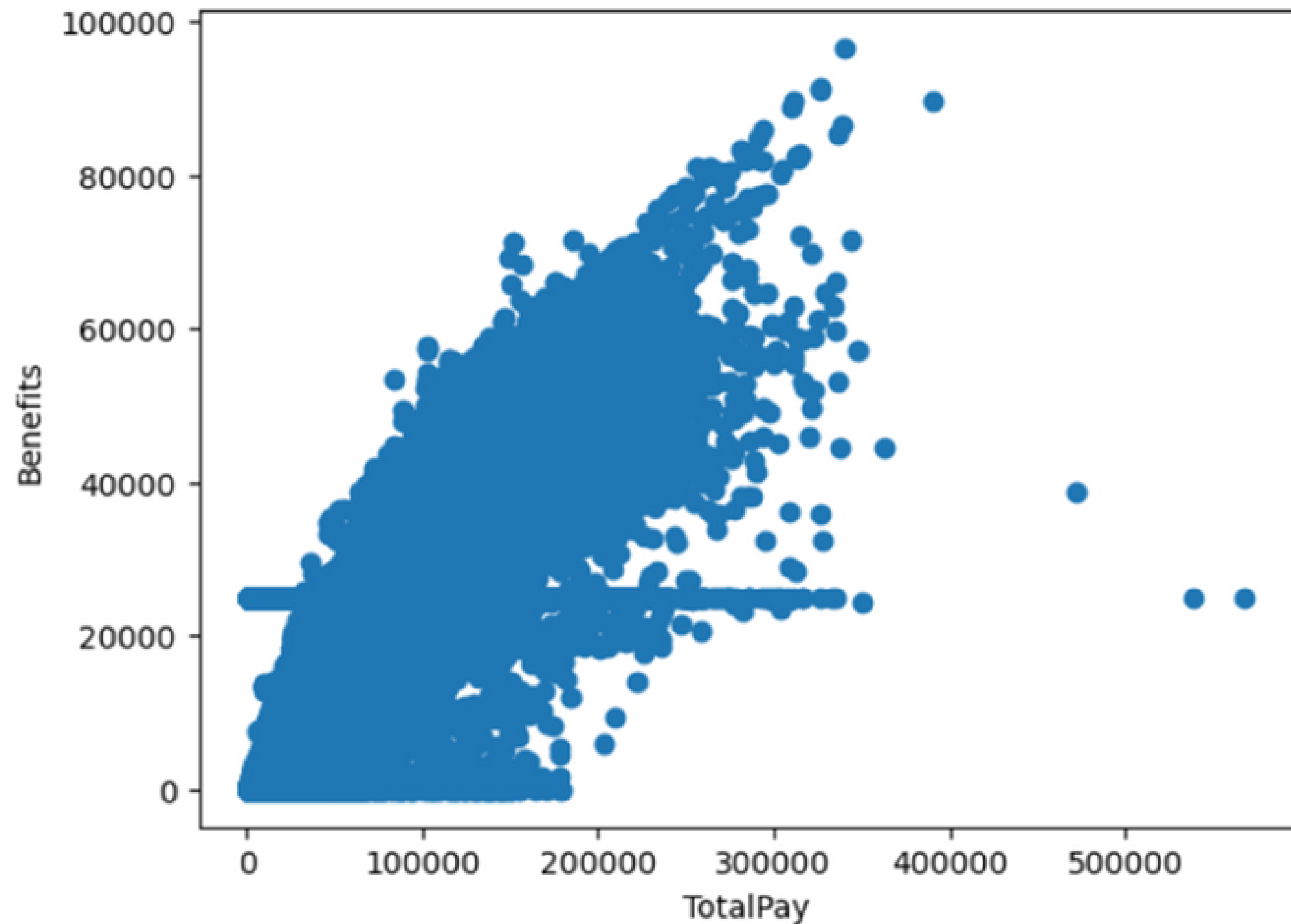
Simple Correlation Analysis

Correlation between TotalPay and TotalPayBenefits



Simple Correlation Analysis

Correlation between TotalPay and Benefits



THANK YOU

FOR YOUR ATTENTION