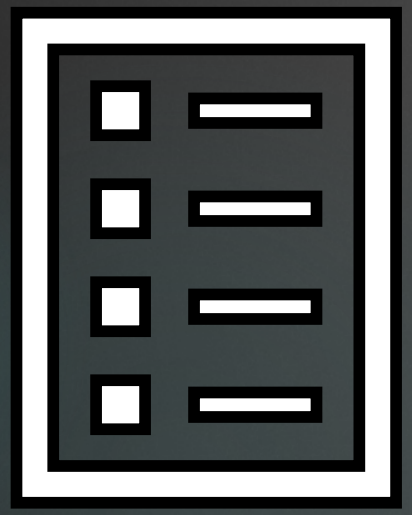
SALARY ANALYSIS REPORTS

FOR SAN FRANCISCO EMPLOYEES

Prepared By: Varun Malhotra

CONTENTS



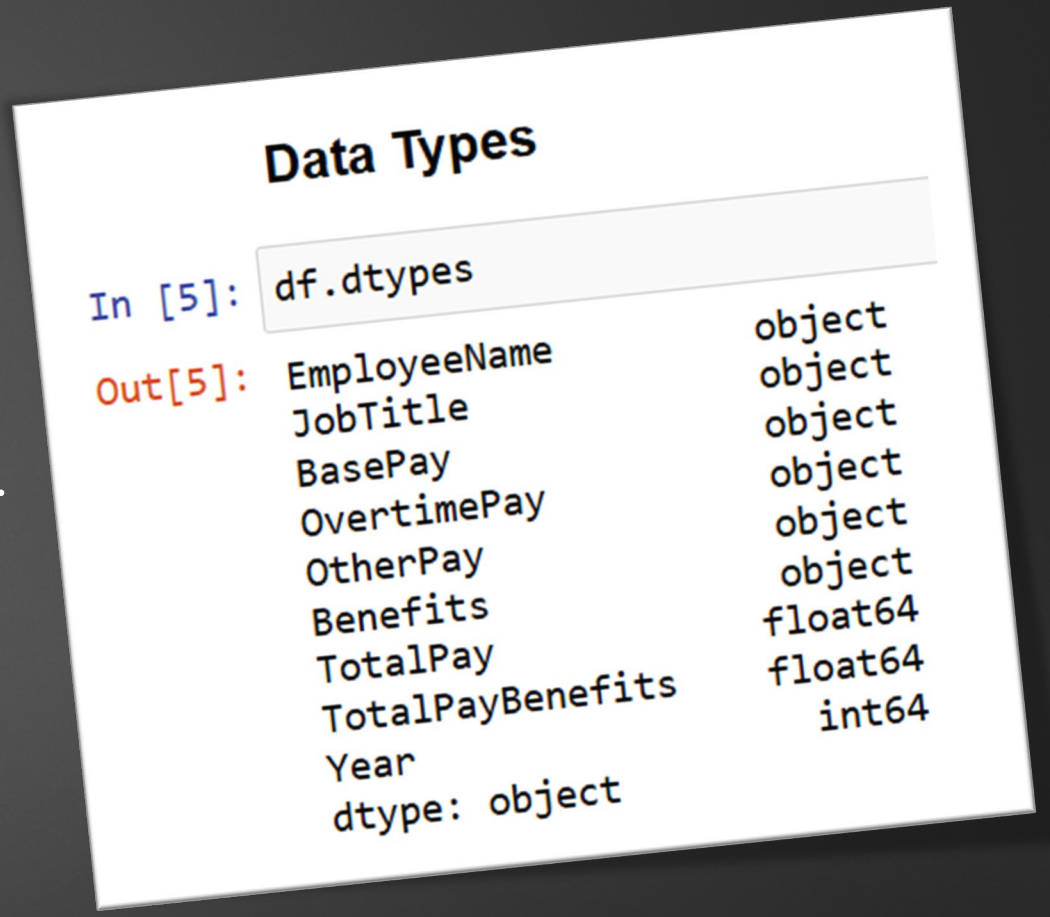
Objective

- The project aims to analyze employee compensation data, including Base Pay, Overtime Pay, Other Pay, Benefits, and their relation to Total Pay and Total Pay Benefits. This is achieved through **Exploratory Data Analysis (EDA)** and **Visualization** using **Python** in Jupyter Notebook and **Microsoft Power BI Desktop**.

Data Overview

Columns in the dataset:

- **EmployeeName**: Name of the employee.
- **JobTitle**: Title of the job.
- **BasePay**: Base salary pay.
- **OvertimePay**: Pay for overtime work.
- **OtherPay**: Any other types of compensation.
- **Benefits**: Benefits provided to the employee.
- **TotalPay**: The total pay without benefits.
- **TotalPayBenefits**: Total pay with benefits included.
- **Year**: The year of the payroll record.



Tools & Technologies Used

Data Cleaning

Handling Null Values

```
In [13]: df.isnull().sum().sum()
Out[13]: 608
```

Typecasting

```
In [8]: # Converting required columns from 'object' to 'numeric' data type
        df['BasePay'] = pd.to_numeric(df['BasePay'], errors = 'coerce')
        df['OvertimePay'] = pd.to_numeric(df['OvertimePay'], errors = 'coerce')
        df['OtherPay'] = pd.to_numeric(df['OtherPay'], errors = 'coerce')
```

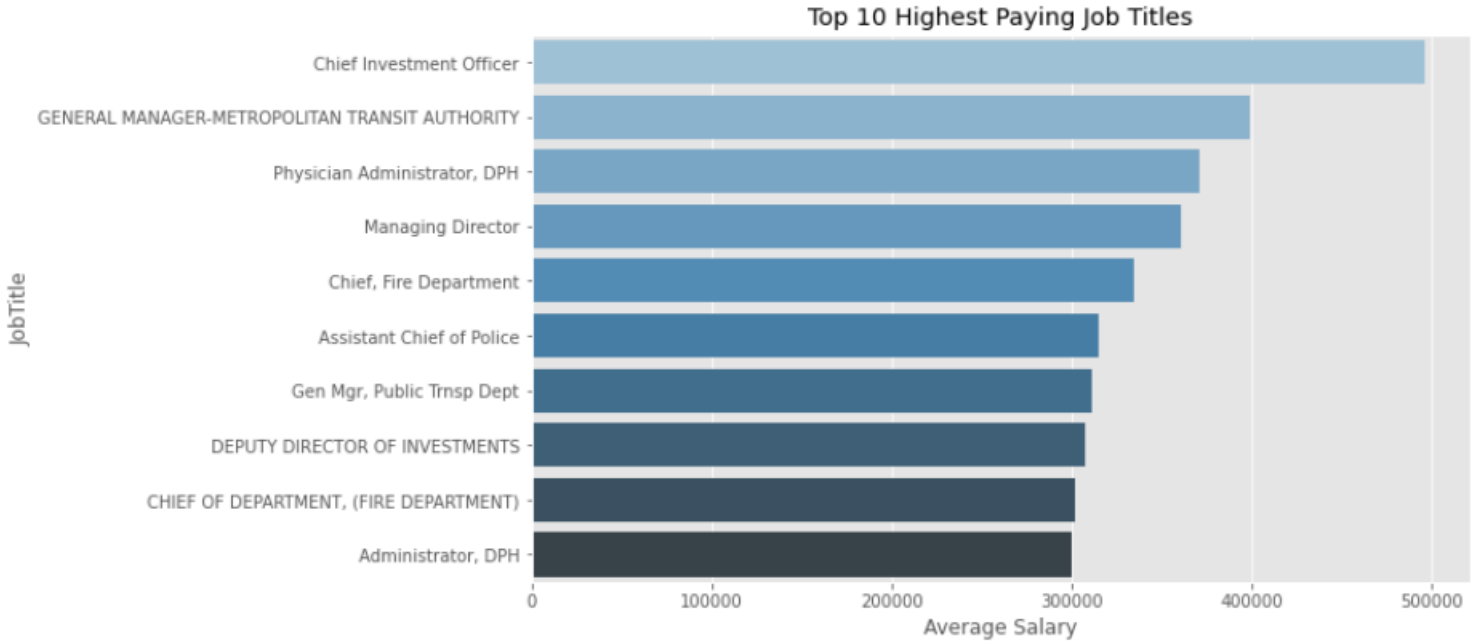
```
In [15]: df['BasePay'].fillna(df['BasePay'].mean(), inplace=True)

In [17]: df=df.dropna(subset=['OvertimePay','OtherPay'])

In [19]: df.isnull().sum().sum()
Out[19]: 0
```

```
In [10]: df.dtypes
Out[10]: EmployeeName        object
         JobTitle            object
         BasePay            float64
         OvertimePay        float64
         OtherPay           float64
         Benefits            object
         TotalPay           float64
         TotalPayBenefits   float64
         Year                 int64
         dtype: object
```

Exploratory Data Analysis

Highest paing Job Titles

In [27]:
```
# Group by job title and get the mean TotalPay
job_salary=df.groupby('JobTitle')['TotalPay'].mean().sort_values(ascending =False).head(10)

# Plot
plt.figure(figsize=(10,6))
sns.barplot(x=job_salary.values, y=job_salary.index, palette='Blues_d')
plt.title('Top 10 Highest Paying Job Titles')
plt.xlabel('Average Salary')
plt.show()
```

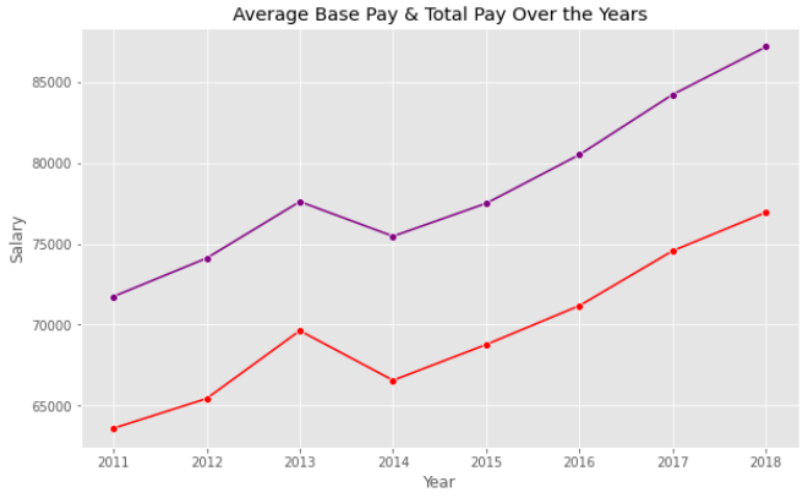


Line Charts

Line Chart Showing

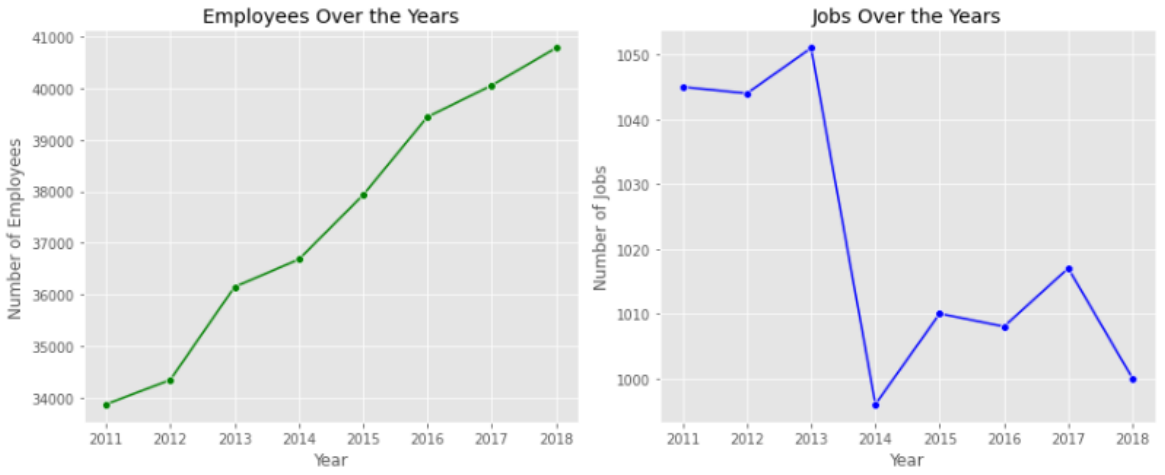
1. Average Pay Over Years

In [28]:
```
# Group by Year and calculate mean total pay
pay_over_years = df.groupby('Year')['TotalPay'].mean()
basepay_over_years = df.groupby('Year')['BasePay'].mean()

# Plot
plt.figure(figsize=(10,6))
sns.lineplot(x=pay_over_years.index, y=pay_over_years.values, marker='o', color='purple')
sns.lineplot(x=basepay_over_years.index, y=basepay_over_years.values, marker='o', color='red')
plt.title('Average Base Pay & Total Pay Over the Years')
plt.xlabel('Year')
plt.ylabel('Salary')
plt.show()
```



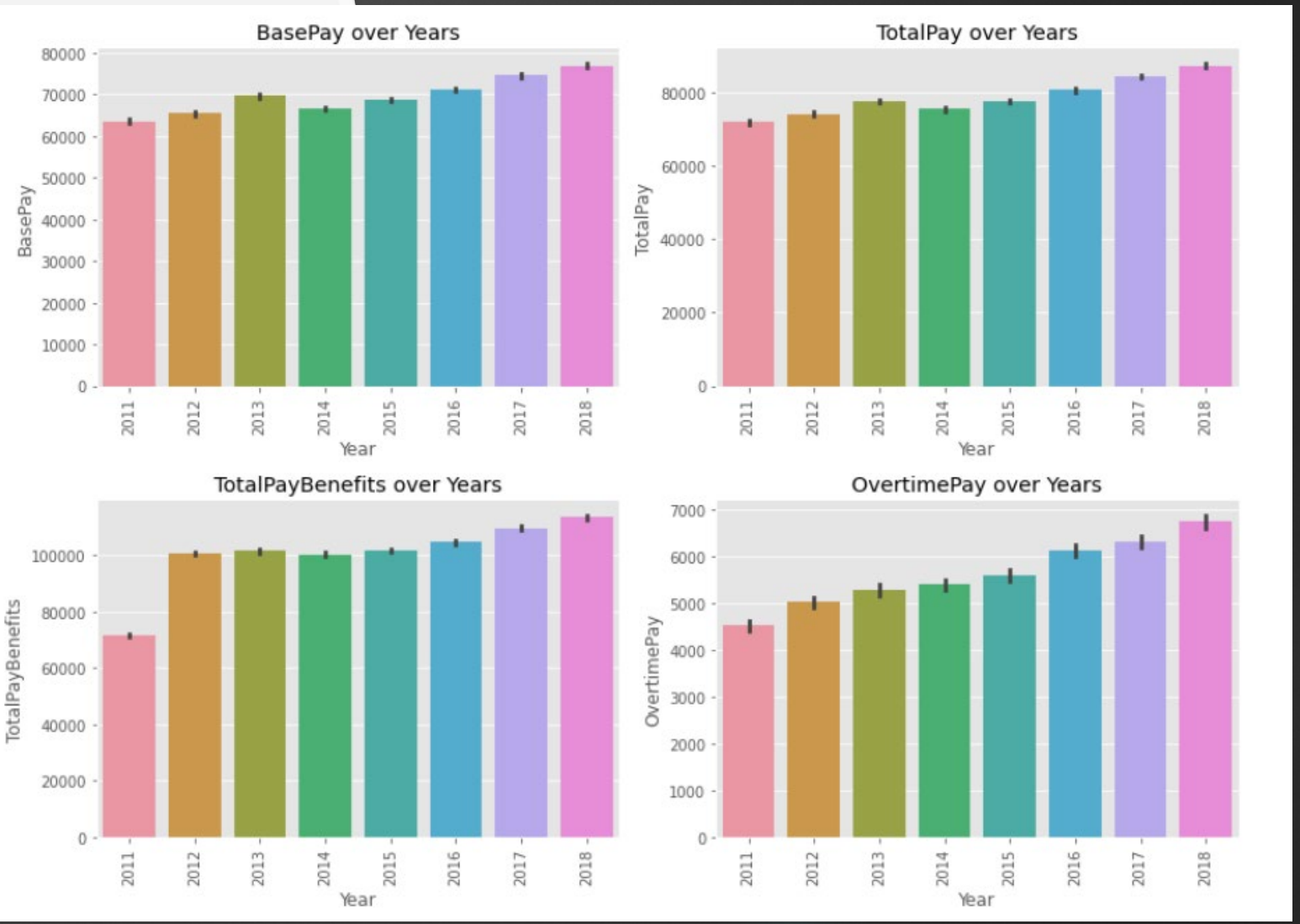
2. Number of employees and jobs over the Years

In [29]:
```
# Calculate number of employees and jobs group by Year
employee_over_years = df.groupby('Year')['EmployeeName'].nunique()
job_over_years = df.groupby('Year')['JobTitle'].nunique()

# Plot linechart for Number of employees and jobs over the Years
plt.figure(figsize=(12,5))

plt.subplot(1,2,1)
sns.lineplot(x=employee_over_years.index, y=employee_over_years.values, marker='o', color='green')
plt.title('Employees Over the Years')
plt.xlabel('Year')
plt.ylabel('Number of Employees')

plt.subplot(1,2,2)
sns.lineplot(x=job_over_years.index, y=job_over_years.values, marker='o', color='blue')
plt.title('Jobs Over the Years')
plt.xlabel('Year')
plt.ylabel('Number of Jobs')

plt.tight_layout()
plt.show()
```



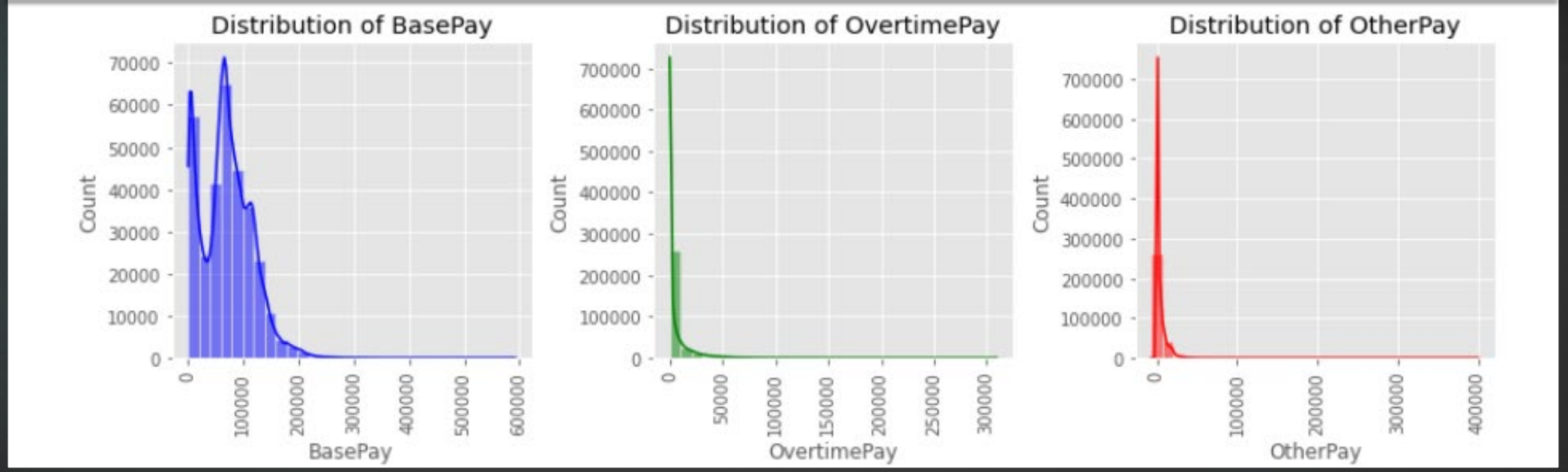
Bar Charts

```
In [30]: # Plot barcharts for BasePay, TotalPay, TotalPayBenefits, OvertimePay, and OtherPay
         plt.figure(figsize=(12,13))

         plt.subplot(3,2,1)
         sns.barplot(data=df, x='Year', y='BasePay')
         plt.xticks(rotation='vertical')
         plt.title('BasePay over Years')

         plt.subplot(3,2,2)
         sns.barplot(data=df, x='Year', y='TotalPay')
         plt.xticks(rotation='vertical')
         plt.title('TotalPay over Years')

         plt.subplot(3,2,3)
         sns.barplot(data=df, x='Year', y='TotalPayBenefits')
         plt.xticks(rotation='vertical')
         plt.title('TotalPayBenefits over Years')

         plt.subplot(3,2,4)
         sns.barplot(data=df, x='Year', y='OvertimePay')
         plt.xticks(rotation='vertical')
         plt.title('OvertimePay over Years')

         plt.subplot(3,2,5)
         sns.barplot(data=df, x='Year', y='OtherPay')
         plt.xticks(rotation='vertical')
         plt.title('OtherPay over Years')

         plt.tight_layout()
         plt.show()
```



Graphs Showing distribution of payments

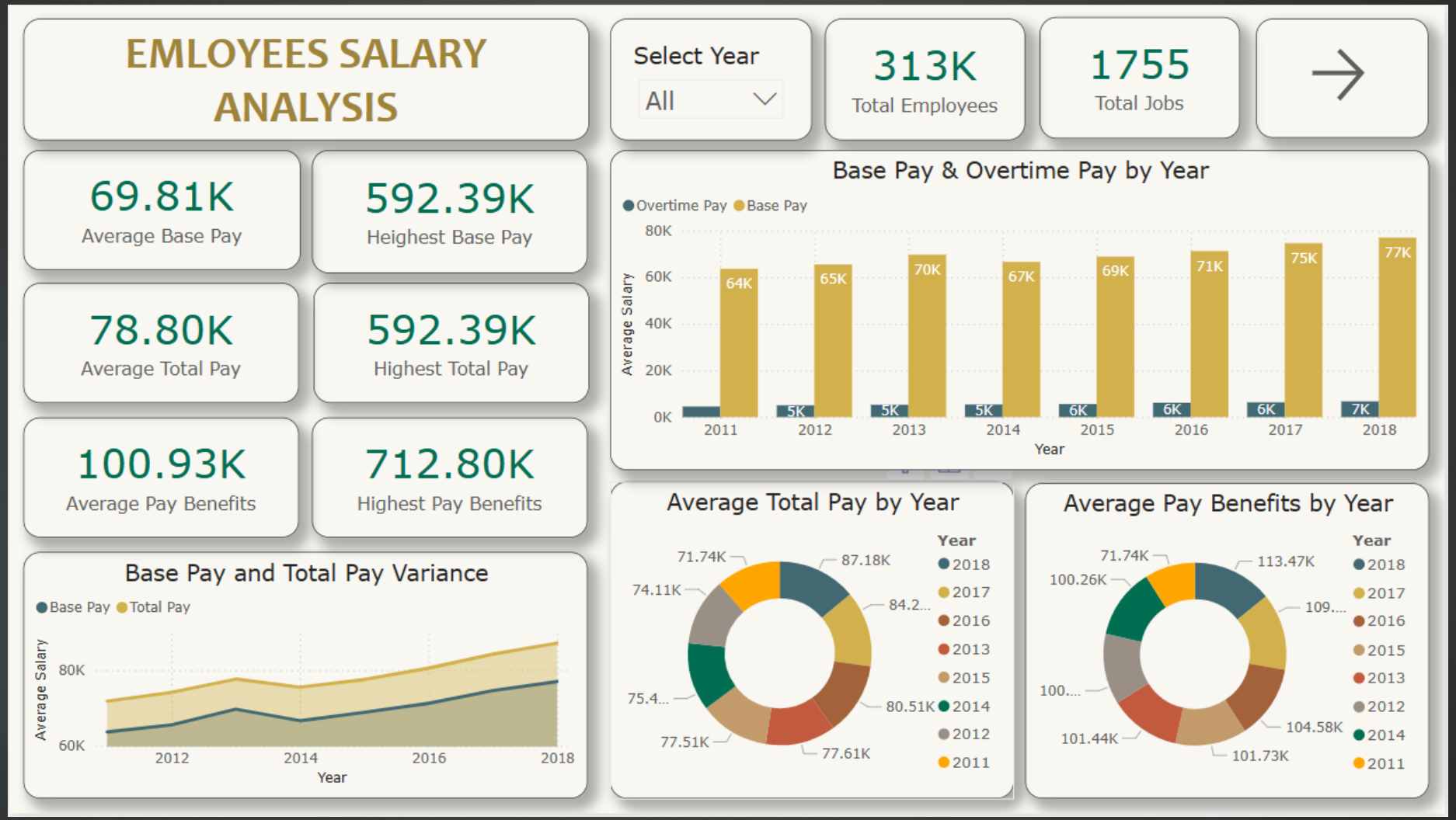
```
In [31]: # Plot histograms for BasePay, OvertimePay, and OtherPay
         plt.figure(figsize=(12,4))

         plt.subplot(1,3,1)
         sns.histplot(df['BasePay'], bins=30, kde=True, color='blue')
         plt.xticks(rotation='vertical')
         plt.title('Distribution of BasePay')

         plt.subplot(1,3,2)
         sns.histplot(df['OvertimePay'], bins=30, kde=True, color='green')
         plt.xticks(rotation='vertical')
         plt.title('Distribution of OvertimePay')

         plt.subplot(1,3,3)
         sns.histplot(df['OtherPay'], bins=30, kde=True, color='red')
         plt.xticks(rotation='vertical')
         plt.title('Distribution of OtherPay')

         plt.tight_layout()
         plt.show()
```



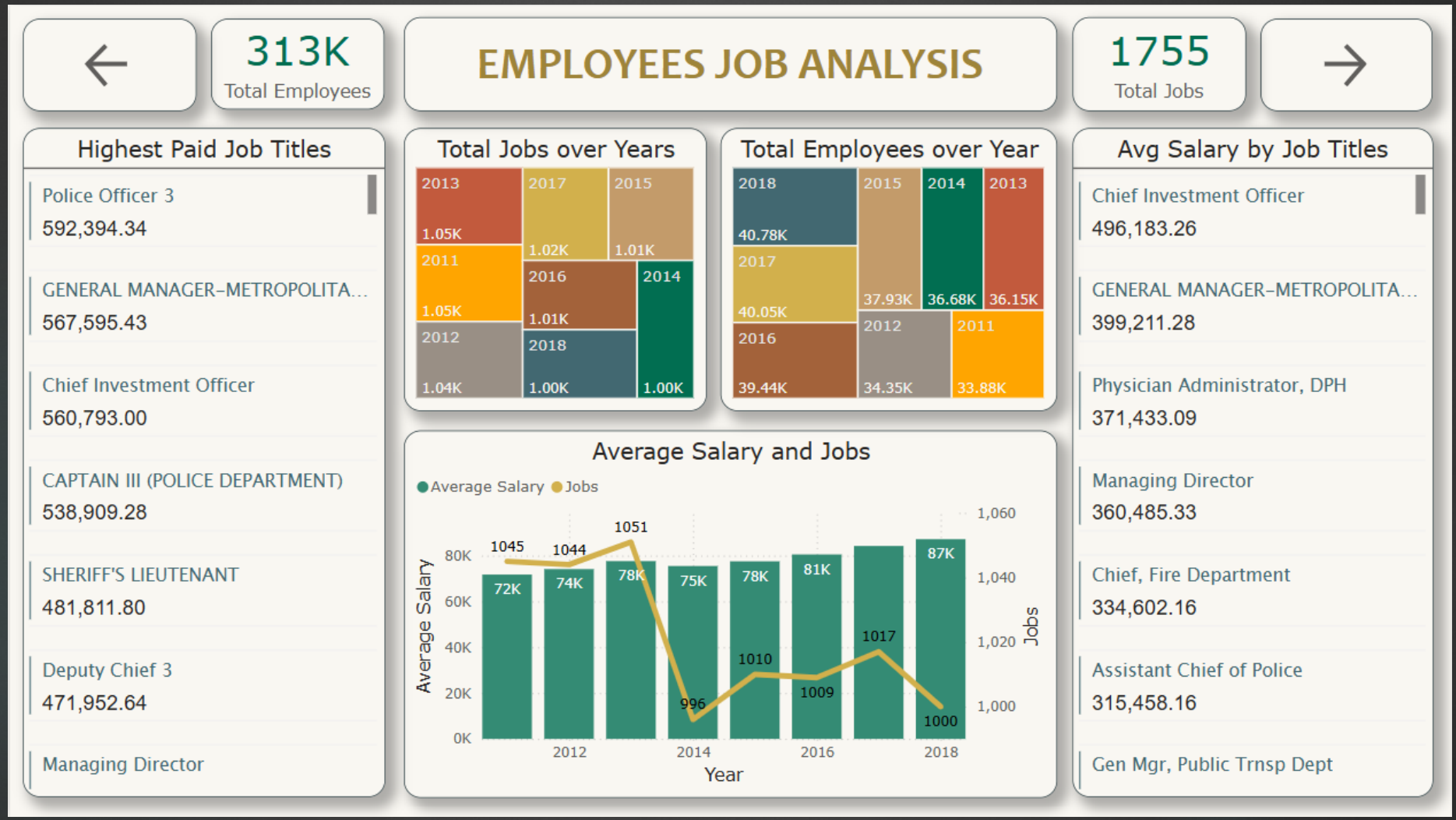
Pie Charts

```
In [45]: plt.figure(figsize=(15,3))

         plt.subplot(1,2,1)
         tmp = df.sort_values(by=['TotalPay'], ascending = False)
         tmp = tmp[['JobTitle','TotalPay']].head()
         plt.pie(tmp.TotalPay, labels = tmp.JobTitle, autopct='%1.1f%%', explode=[0.1,0,0,0,0])
         plt.title('Total Pay by Job Title')

         plt.subplot(1,2,2)
         tmp = df.sort_values(by=['TotalPayBenefits'], ascending = False)
         tmp = tmp[['JobTitle','TotalPayBenefits']].head()
         plt.pie(tmp.TotalPayBenefits, labels = tmp.JobTitle, autopct='%1.1f%%', explode=[0.1,0,0,0,0])
         plt.title('Total Pay Benefits by Job Title')

         plt.tight_layout()
         plt.show()
```

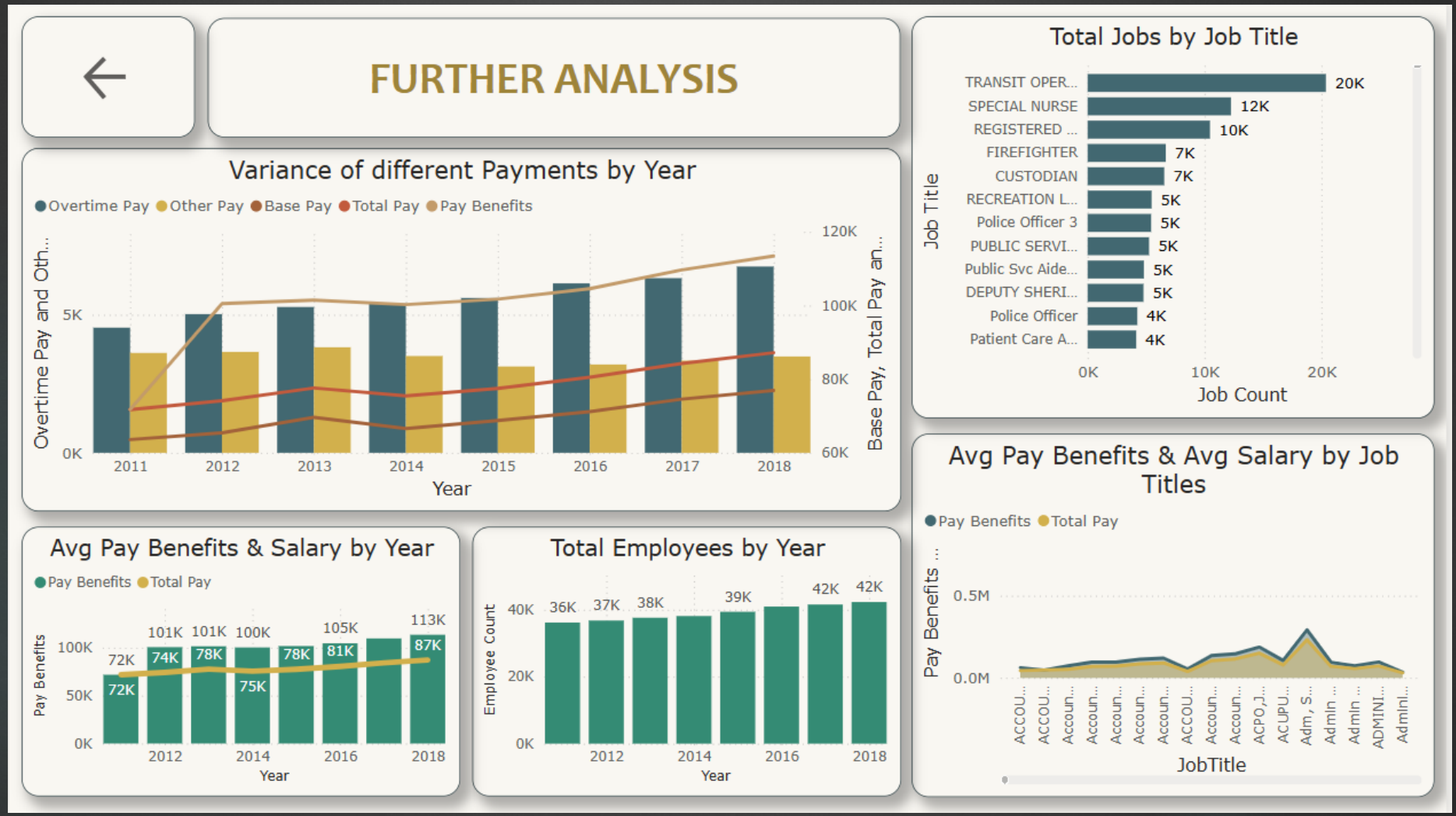
Dashboard Components



Dashboard Components



Dashboard Components



Key Metrics Summary

Salary Overview:

- **Average Base Pay:** $69.81K
- **Highest Base Pay:** $592.39K
- **Average Total Pay:** $78.80K
- **Highest Total Pay:** $592.39K
- **Average Pay Benefits:** $100.93K
- **Highest Pay Benefits:** $712.80K

Workforce Overview:

- **Total Employees:** 313K
- **Total Job Titles:** 1,755

Trend Analysis

Base Pay & Overtime Trends

- From 2011 to 2018, **Base Pay** consistently increased from $64K to $77K.
- **Overtime Pay** also showed a gradual increase, reflecting growing demand or workload.

Year-wise Variance

- The line and bar charts show a **steady upward trend in all salary components**, especially after 2014.
- Benefits and overtime form a growing share of total pay, suggesting either better benefit programs or greater reliance on overtime.

Total Pay & Pay Benefits

- **Total Pay** rose from $72K (2011) to $87K (2018).
- **Pay Benefits** climbed significantly from $72K to $113K, indicating a shift toward benefits-focused compensation.

Highest Paid Titles

- **Police Officer 3:** $592,394
- **General Manager–Metropolitan:** $567,595
- **Chief Investment Officer:** $560,793
- **Captain III (Police Department):** $538,909
- High earners are predominantly from **law enforcement and executive roles**.

Trend Analysis

Average Salary by Title

- **Chief Investment Officer:** $496,183 (highest average)
- **Assistant Chief of Police:** $315,458
- Reflects that **administrative and public safety roles** dominate top average salaries.

Total Jobs by Year

- Job counts have remained relatively stable (~1,000 per year).
- Minor fluctuations suggest consistent hiring practices or job reclassifications.

Employee Count Over Years

- Number of employees increased from **~34K in 2011** to **~42K in 2018**.
- Growth implies workforce expansion or regular onboarding.

Jobs by Title

- **Transit Operator:** 20K jobs
- **Special Nurse:** 12K jobs
- **Registered Nurse:** 10K jobs
- Job volume indicates **healthcare and transit are the most populous sectors**.

Trend Analysis

Additional Insights:

Average Pay Benefits vs Avg Salary by Job

- The chart shows job titles with **disproportionately higher benefits** compared to salary.
- Indicates potential areas of **cost optimization or restructuring**.

Component Variance

- From 2011 to 2018, **overtime, pay benefits, and base pay** showed consistent growth.
- Increasing **benefit load** suggests a strategic shift or legal compliance changes.

Conclusion

- This analysis provides a detailed view of the compensation landscape across years and job titles. With rising salaries, expanding benefits, and consistent job volumes, the data supports informed decision-making for HR planning, salary benchmarking, and organizational budgeting.

THANK YOU