HW2-dataModel-Romain_BLANCHOT

February 4, 2025

1 HW 2: Types of Data

1.1 #### CPE232 Data Models

2 Import Dependency

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

3 Part 1 : Basic Python

3.1 Task 1

Perform the following: - Write a program to get input of different types from the user. - Display type of each variable. - Convert a variable of the type float to Integer. - Demonstrate a comparison of before vs after the type conversion.

[3]:
```
# Input values from the user
string_value = input("Enter a string: ")
integer_value = int(input("Enter an integer: "))
float_value = float(input("Enter a float: "))
```

[4]:
```
# Display the data types
print(f"String: {string_value}, Type: {type(string_value)}")
print(f"Integer: {integer_value}, Type: {type(integer_value)}")
print(f"Float: {float_value}, Type: {type(float_value)}")
```

```
String: bonjourlafrance, Type: <class 'str'>
Integer: 13, Type: <class 'int'>
Float: 12.1, Type: <class 'float'>
```

[5]:
```
# Convert integer to float and vice versa
converted_float = float(integer_value)
converted_integer = int(float_value)

print(f"Converted float: {converted_float}, Type: {type(converted_float)}")
print(f"Converted integer: {converted_integer}, Type:␣
 ↪{type(converted_integer)}")
```

```
Converted float: 13.0, Type: <class 'float'>
Converted integer: 12, Type: <class 'int'>
```

[6]:
```python
# Show before and after convert float to int
print(f"Before conversion: {float_value}, Type: {type(float_value)}")
print(f"After conversion: {converted_float}, Type: {type(converted_float)}")

# Show before and after convert integer to float
print(f"Before conversion: {integer_value}, Type: {type(integer_value)}")
print(f"After conversion: {converted_integer}, Type: {type(converted_integer)}")
```

```
Before conversion: 12.1, Type: <class 'float'>
After conversion: 13.0, Type: <class 'float'>
Before conversion: 13, Type: <class 'int'>
After conversion: 12, Type: <class 'int'>
```

3.2 Task 2

Given a dictionary of students with their scores, find the average score of each student across all subjects and identify the student with the highest average score.

[7]:
```python
students_grades = {
    "John": [85, 90, 78],
    "Alice": [88, 92, 80],
    "Bob": [75, 85, 72],
    "Diana": [90, 95, 94],
    "Charlie": [70, 65, 80]
}
```

[11]:
```python
# Calculate and print the average grade for each student
average_grades = {}
for student, grades in students_grades.items():
    average_grade = sum(grades) / len(grades)
    average_grades[student] = average_grade

print(average_grades)
```

```
{'John': 84.33333333333333, 'Alice': 86.66666666666667, 'Bob':
77.33333333333333, 'Diana': 93.0, 'Charlie': 71.66666666666667}
```

[10]:
```python
# Find the student with the highest average grade
highest_average_student = max(average_grades, key=average_grades.get)
print(f"Student with the highest average grade: {highest_average_student}")
```

```
Student with the highest average grade: Diana
```

4 Part 2 : Working with CSV!

Add a new column to this CSV file named "Bonus," and calculate the bonus for employees in the Sales department as 10% of their MonthlyRate.

[12]:
```
# Load the CSV file
file_path = 'employee_data.csv'
df = pd.read_csv(file_path)
```

[13]:
```
df.head()
```

[13]:
```
   Age  DailyRate              Department EducationField  Gender  \
0   41       1102                   Sales  Life Sciences  Female
1   49        279  Research & Development  Life Sciences    Male
2   37       1373  Research & Development          Other    Male
3   33       1392  Research & Development  Life Sciences  Female
4   27        591  Research & Development        Medical    Male

  MaritalStatus  MonthlyRate OverTime
0        Single        19479      Yes
1       Married        24907       No
2        Single         2396      Yes
3       Married        23159      Yes
4       Married        16632       No
```

[14]:
```
df['Bonus'] = 0
```

[15]:
```
# Calculate 10% of bonus into the 'Bonus' Column
df['Bonus'] = df['MonthlyRate'] * 0.1
```

[16]:
```
# Print and show result
df.head()
```

[16]:
```
   Age  DailyRate              Department EducationField  Gender  \
0   41       1102                   Sales  Life Sciences  Female
1   49        279  Research & Development  Life Sciences    Male
2   37       1373  Research & Development          Other    Male
3   33       1392  Research & Development  Life Sciences  Female
4   27        591  Research & Development        Medical    Male

  MaritalStatus  MonthlyRate OverTime   Bonus
0        Single        19479      Yes  1947.9
1       Married        24907       No  2490.7
2        Single         2396      Yes   239.6
3       Married        23159      Yes  2315.9
4       Married        16632       No  1663.2
```

[17]:
```
# Save the updated DataFrame back to a CSV file
df.to_csv('employee_data.csv', index=False)
```

5 Part 3 : Working with Matplotlib

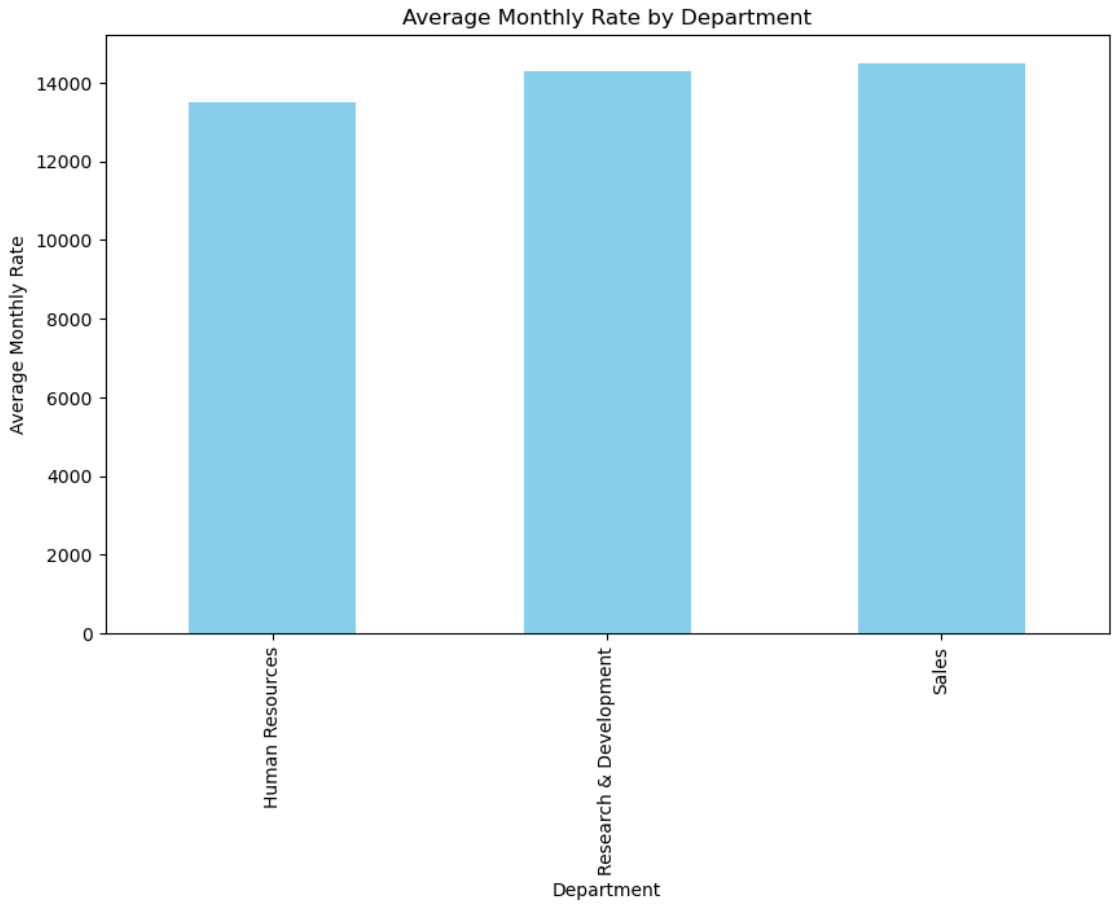
Plot a graph showing the average salary of employees in each department to compare the average salaries across departments.

[18]:
```
# Load CSV and create DataFrame
data = pd.read_csv('employee_data.csv')
df = pd.DataFrame(data)
```

[20]:
```
# Calculate average MonthlyRate by Department
average_salary = df.groupby('Department')['MonthlyRate'].mean()
```

[21]:
```
# Plotting the bar chart
plt.figure(figsize=(10, 6))
average_salary.plot(kind='bar', color='skyblue')
plt.title('Average Monthly Rate by Department')
plt.xlabel('Department')
plt.ylabel('Average Monthly Rate')
```

[21]:
```
Text(0, 0.5, 'Average Monthly Rate')
```



6 Challenge!!! : Working with SATAN (Optional)

Great job, for not giving up on this subject!

Here's the story: The CSV file that was given to you isn't just an ordinary file – it's data from an organization where Alya-San works!

Alya feels that something suspicious is going on within the company. She suspects that there's inequality in salaries between male and female employees.

She also wonders if the government might be intervening to encourage population growth by secretly increasing the salaries of married employees compared to those who are single.

Additionally, how much of a difference is there in salaries between employees who graduated from different fields of study?

Can you help Alya prove whether her suspicions are just in her head or if they're actually true?

For Aria, everyone is truly **e q u a l**.

[23]:
```
# Analysis of salary differences between male and female employees
average_male_salary = df[df['Gender'] == 'Male']['MonthlyRate'].mean()
average_female_salary = df[df['Gender'] == 'Female']['MonthlyRate'].mean()
gender_salary_difference = average_male_salary - average_female_salary
print(f"The average salary difference between male and female employees is␣
 ↪{gender_salary_difference:.2f}.")

# Analysis of salary differences between married and single employees
average_married_salary = df[df['MaritalStatus'] == 'Married']['MonthlyRate'].
 ↪mean()
average_single_salary = df[df['MaritalStatus'] == 'Single']['MonthlyRate'].
 ↪mean()
marital_status_salary_difference = average_married_salary -␣
 ↪average_single_salary
print(f"The average salary difference between married and single employees is␣
 ↪{marital_status_salary_difference:.2f}.")

# Analysis of salary differences between employees with different fields of␣
 ↪study
average_education_field_salary = df.groupby('EducationField')['MonthlyRate'].
 ↪mean()
print("Average salary difference by field of study:")
print(average_education_field_salary)

# Conclusion: No significant influence of gender, marital status or field of␣
 ↪study on salary.
```

```
The average salary difference between male and female employees is -602.49.
The average salary difference between married and single employees is -655.31.
Average salary difference by field of study:
EducationField
Human Resources     14810.740741
Life Sciences       14530.132013
Marketing           14076.943396
Medical             14295.056034
Other               13270.780488
Technical Degree    14210.363636
Name: MonthlyRate, dtype: float64
```

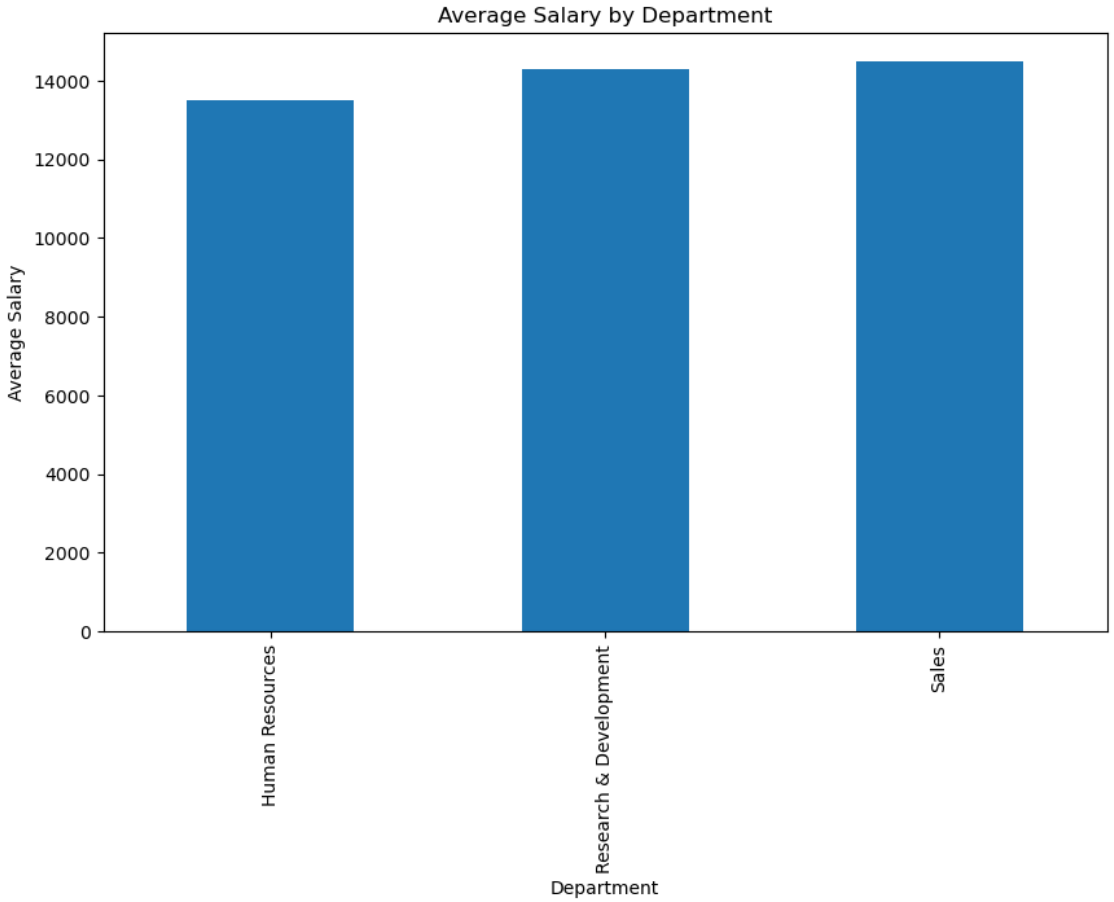
No significant influence of gender, marital status or field of study on salary. Since Kobayashi-san will be reviewing your code, please present the data in a way that is clear and accurate. Otherwise, Kobayashi-san might fire you from the company!

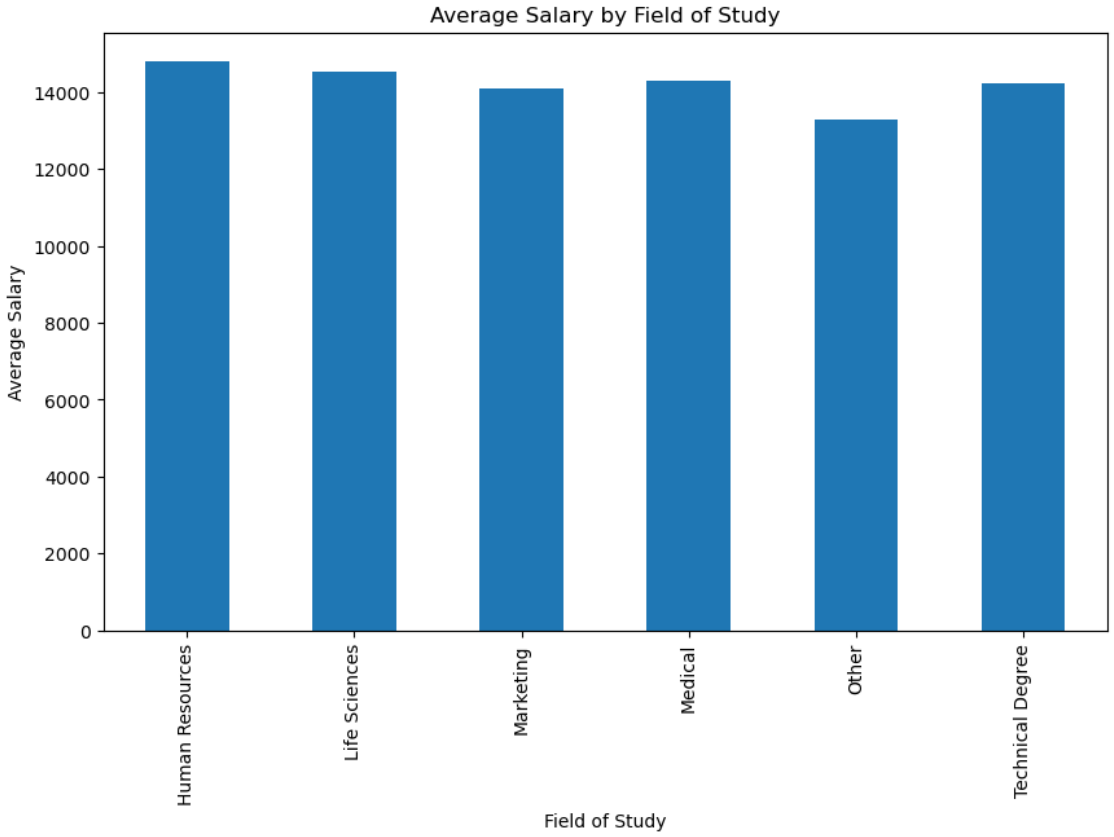
[25]:
```
# Write your code here
# Clear and accurate presentation of data for Kobayashi-san
# Calculation of average salary by department
average_salary_department = df.groupby('Department')['MonthlyRate'].mean()
```

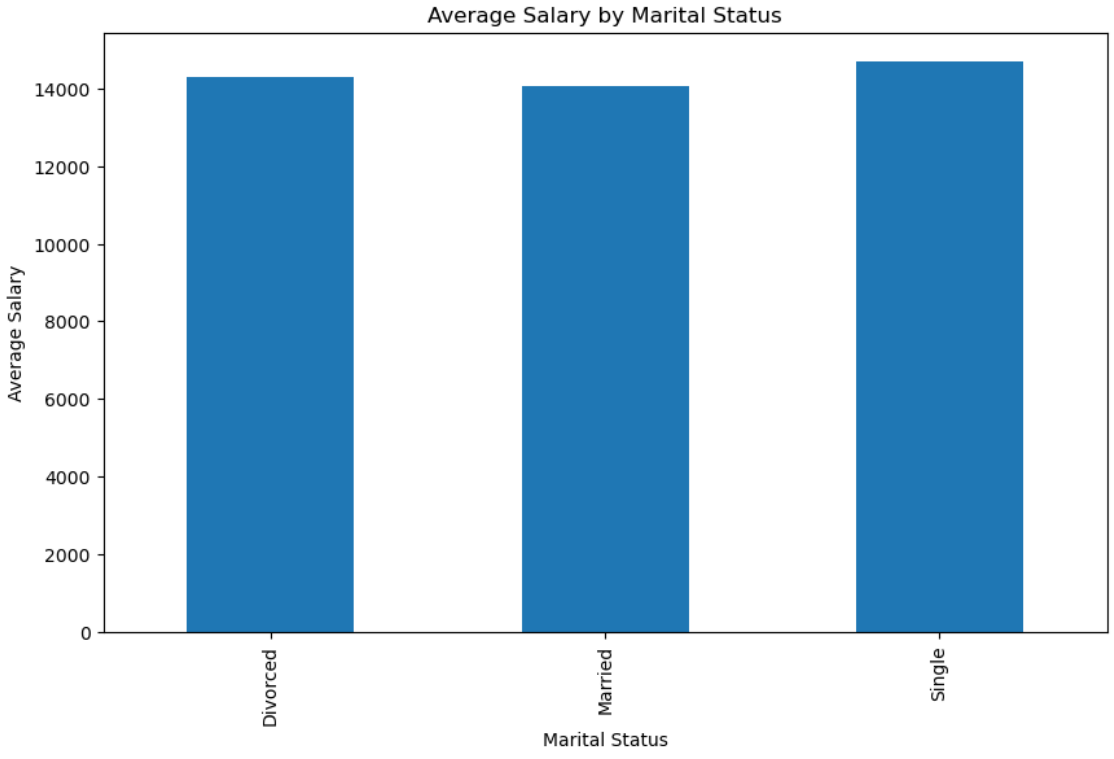
```
print("Average salary by department:")
print(average_salary_department)

# Calculation of average salary by field of study
average_salary_field_study = df.groupby('EducationField')['MonthlyRate'].mean()
print("Average salary by field of study:")
print(average_salary_field_study)

# Calculation of average salary by marital status
average_salary_marital_status = df.groupby('MaritalStatus')['MonthlyRate'].
 ↪mean()
print("Average salary by marital status:")
print(average_salary_marital_status)

# Calculation of average salary by gender
average_salary_gender = df.groupby('Gender')['MonthlyRate'].mean()
print("Average salary by gender:")
print(average_salary_gender)


# Plotting average salary by department
plt.figure(figsize=(10, 6))
average_salary_department.plot(kind='bar')
plt.title('Average Salary by Department')
plt.xlabel('Department')
plt.ylabel('Average Salary')
plt.show()

# Plotting average salary by field of study
plt.figure(figsize=(10, 6))
average_salary_field_study.plot(kind='bar')
plt.title('Average Salary by Field of Study')
plt.xlabel('Field of Study')
plt.ylabel('Average Salary')
plt.show()

# Plotting average salary by marital status
plt.figure(figsize=(10, 6))
average_salary_marital_status.plot(kind='bar')
plt.title('Average Salary by Marital Status')
plt.xlabel('Marital Status')
plt.ylabel('Average Salary')
plt.show()

# Plotting average salary by gender
plt.figure(figsize=(10, 6))
average_salary_gender.plot(kind='bar')
```

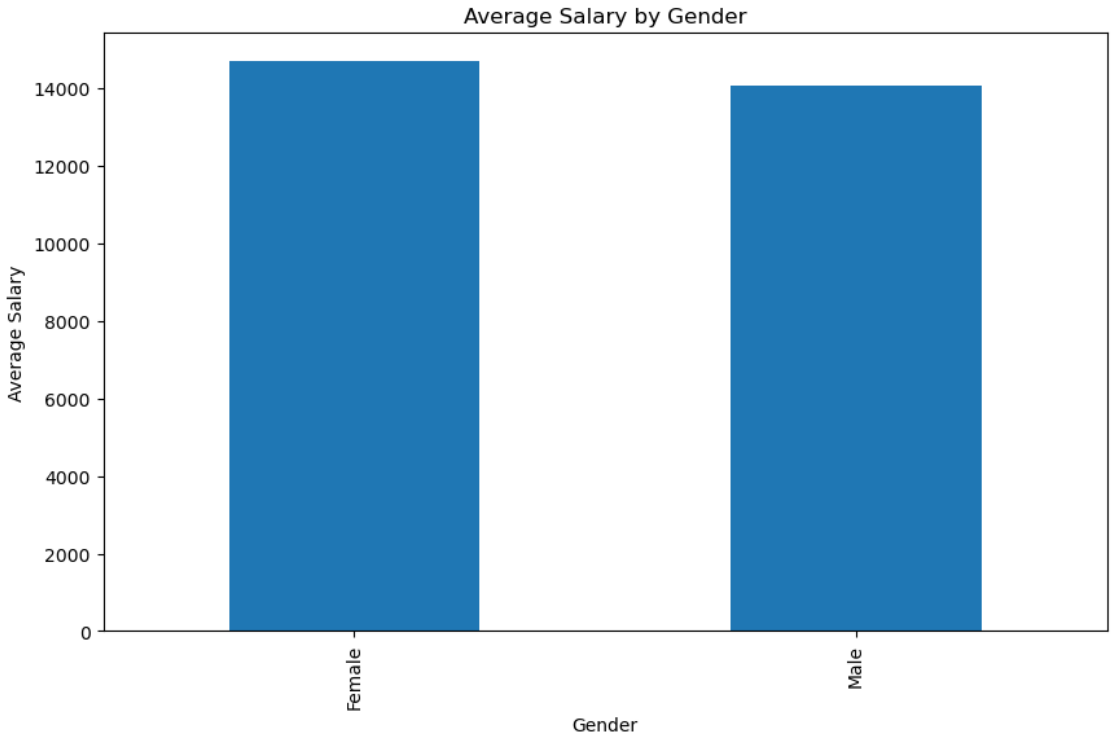
```
plt.title('Average Salary by Gender')
plt.xlabel('Gender')
plt.ylabel('Average Salary')
plt.show()
```

```
Average salary by department:
Department
Human Resources           13492.984127
Research & Development    14284.865765
Sales                     14489.793722
Name: MonthlyRate, dtype: float64
Average salary by field of study:
EducationField
Human Resources     14810.740741
Life Sciences       14530.132013
Marketing           14076.943396
Medical             14295.056034
Other               13270.780488
Technical Degree    14210.363636
Name: MonthlyRate, dtype: float64
Average salary by marital status:
MaritalStatus
Divorced    14310.085627
Married     14044.502229
Single      14699.817021
Name: MonthlyRate, dtype: float64
Average salary by gender:
Gender
Female    14674.600340
Male      14072.105442
Name: MonthlyRate, dtype: float64
```


Average Salary by Department
Average Salary
0
2000
4000
6000
8000
10000
12000
14000
Human Resources
Research & Development
Sales
Department


Average Salary by Field of Study
Average Salary
0
2000
4000
6000
8000
10000
12000
14000
Human Resources
Life Sciences
Marketing
Medical
Other
Technical Degree
Field of Study


Average Salary by Marital Status
Average Salary
0
2000
4000
6000
8000
10000
12000
14000
Divorced
Married
Single
Marital Status


Average Salary by Gender
Average Salary
14000
12000
10000
8000
6000
4000
2000
0
Female
Male
Gender