

HW 2: Types of Data

CPE232 Data Models

Import Dependency

```
In [4]: # pip install seaborn
```

```
In [22]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Part 1 : Basic Python

Task 1

เขียนโปรแกรมสำหรับใส่ค่า Input แต่ละ type ที่ต่างกัน ทำการแสดง type ของแต่ละตัว และทดลองแปลงค่า Float ให้กลายเป็น int พร้อมแสดง Before After

```
In [6]: # 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: "))
```

```
Enter a string: Kiki
Enter an integer: 12
Enter a float: 8.6
```

```
In [7]: # Display the data types
print(type(string_value))
print(type(integer_value))
print(type(float_value))
```

```
<class 'str'>
<class 'int'>
<class 'float'>
```

```
In [8]: # Convert integer to float and vice versa
after1 = float(integer_value)
after2 = int(float_value)
```

```
In [9]: # Show before and after convert float to int
print("Before convert integer to float:", integer_value)
print("After convert integer to float:", after1)
print("Before convert float to integer:", float_value)
print("After convert float to integer:", after2)
```

Before convert integer to float: 12
 After convert integer to float: 12.0
 Before convert float to integer: 8.6
 After convert float to integer: 8

Task 2

มีคะแนนของนักเรียนแต่ละคนให้ในรูปแบบ Dictionary จงหาคะแนนเฉลี่ยของนักเรียนแต่ละคนในทุกวิชาและหาคนที่ได้คะแนนเฉลี่ยสูงสุด

```
In [10]: students_grades = {
    "John": [85, 90, 78],
    "Alice": [88, 92, 80],
    "Bob": [75, 85, 72],
    "Diana": [90, 95, 94],
    "Charlie": [70, 65, 80]
}
```

```
In [11]: # Calculate and print the average grade for each student
listGrade = []
for student, grades in students_grades.items():
    average = sum(grades) / len(grades)
    listGrade.append(average)
    print(f"{student} has an average grade of {average:.2f}")
```

John has an average grade of 84.33
 Alice has an average grade of 86.67
 Bob has an average grade of 77.33
 Diana has an average grade of 93.00
 Charlie has an average grade of 71.67

```
In [12]: # Find the student with the highest average grade
index = listGrade.index(max(listGrade))
student = list(students_grades.keys())[index]
print('The student with the highest average grade is', student + ".")
```

The student with the highest average grade is Diana.

Part 2 : Working with CSV!

จงสร้างคอลัมน์ใหม่ให้กับ CSV ไฟล์นี้ชื่อว่า Bonus และให้ค่านวน Bonus ให้กับพนักงานที่อยู่แผนก Sales เป็นจำนวน 10% จาก MonthlyRate ของคนๆนั้น

```
In [13]: # Load the CSV file
file_path = 'employee_data.csv'
df = pd.read_csv(file_path)
```

```
In [14]: df.head()
```

Out[14]:

	Age	DailyRate	Department	EducationField	Gender	MaritalStatus	MonthlyRate
0	41	1102	Sales	Life Sciences	Female	Single	19479
1	49	279	Research & Development	Life Sciences	Male	Married	24907
2	37	1373	Research & Development	Other	Male	Single	2396
3	33	1392	Research & Development	Life Sciences	Female	Married	23159
4	27	591	Research & Development	Medical	Male	Married	16632

In [15]: `# Add new column call 'Bonus'`
`df['Bonus'] = ''`
`print(df.columns)`

Index(['Age', 'DailyRate', 'Department', 'EducationField', 'Gender',
'MaritalStatus', 'MonthlyRate', 'OverTime', 'Bonus'],
dtype='object')

In [16]: `# Calculate 10% of bonus into the 'Bonus' Column`
`df['Bonus'] = df['MonthlyRate'] * 0.1`

In [17]: `# Print and show result`
`print(df[['MonthlyRate', 'Bonus']].head())`

	MonthlyRate	Bonus
0	19479	1947.9
1	24907	2490.7
2	2396	239.6
3	23159	2315.9
4	16632	1663.2

In [18]: `# Save the updated DataFrame back to a CSV file`
`df.to_csv('employee_data_updated.csv', index=False)`

Part 3 : Working with Matplotlib

จงพล็อตกราฟแสดงค่าเฉลี่ยเงินเดือนของพนักงานในแต่ละ Department เพื่อเปรียบเทียบเงินเดือนเฉลี่ยของแต่ละ Department

In [19]: `# Load CSV and create DataFrame`
`data = pd.read_csv('employee_data.csv')`
`df = pd.DataFrame(data)`
`df`

Out[19]:

	Age	DailyRate	Department	EducationField	Gender	MaritalStatus	MonthlyRate
0	41	1102	Sales	Life Sciences	Female	Single	1947
1	49	279	Research & Development	Life Sciences	Male	Married	2490
2	37	1373	Research & Development	Other	Male	Single	239
3	33	1392	Research & Development	Life Sciences	Female	Married	2315
4	27	591	Research & Development	Medical	Male	Married	1663
...
1465	36	884	Research & Development	Medical	Male	Married	1229
1466	39	613	Research & Development	Medical	Male	Married	2145
1467	27	155	Research & Development	Life Sciences	Male	Married	517
1468	49	1023	Sales	Medical	Male	Married	1324
1469	34	628	Research & Development	Medical	Male	Married	1022

1470 rows × 8 columns



In [20]:

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

Out[20]:

	MonthlyRate
Department	
Human Resources	13492.984127
Research & Development	14284.865765
Sales	14489.793722

dtype: float64

In [43]:

```
# Plotting the bar chart
average_monthly_rate = df.groupby('Department')['MonthlyRate'].mean().reset_index()

sns.set(style="whitegrid")

plt.figure(figsize=(8, 6))
sns.barplot(x='Department', y='MonthlyRate', data=average_monthly_rate, palette=
```

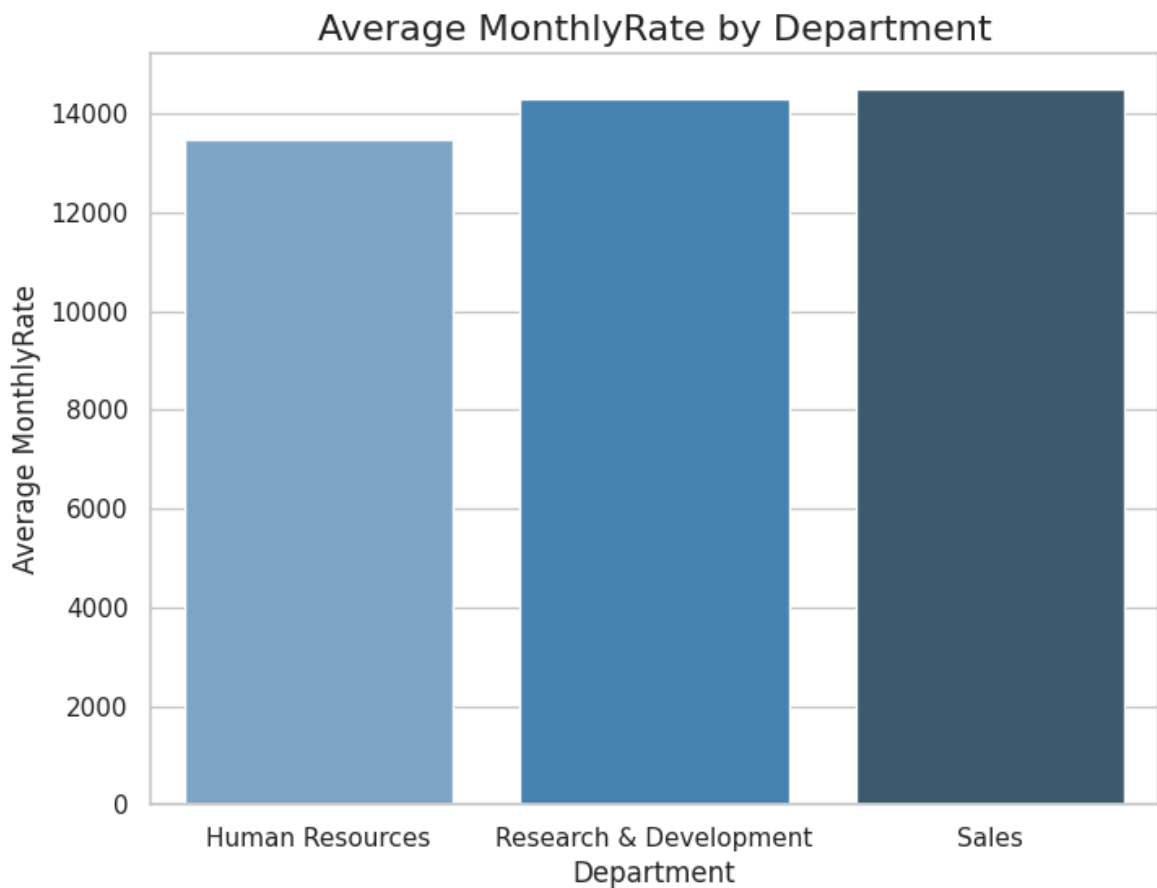
```
plt.title('Average MonthlyRate by Department', fontsize=16)
plt.xlabel('Department', fontsize=12)
plt.ylabel('Average MonthlyRate', fontsize=12)

plt.show()
```

<ipython-input-43-3fa5985cc4c0>:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v 0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='Department', y='MonthlyRate', data=average_monthly_rate, palette='Blues_d')
```



Challenge!!! : Working with SATAN (Optional)

ดีมากเหล่าเด็กๆที่ยังไม่ยอมแพ้ให้กับวิชานี้!



เรื่องมันมีอยู่ว่า จากไฟล์ CSV ที่พี่ไทให้น้องๆไป...มันไม่ใช่ไฟล์ธรรมดา แต่มันเป็นไฟล์ข้อมูลในองค์กรหนึ่งที่ Alya-San นั้นได้ทำงานอยู่



ซึ่งคุณอาเรียรู้สึกว่ามีสิ่งไม่ชอบมาพากลภายในบริษัท ว่ารายได้ของพนักงานชายและหญิงมีความไม่เท่าเทียมกัน

รวมถึงรัฐอาจมีการแทรกแซงเพื่อให้ประชากรภายในประเทศเพิ่มสูงขึ้น รัฐจึงแอบเพิ่มเงินเดือนให้กับผู้ที่แต่งงานอยู่ริเปล่าเมื่อเทียบกับคนโสด

และคนที่จบจากสาขาที่แตกต่างกัน พวกเขาได้เงินเดือนมากกว่ากันมากน้อยแค่ไหน

น้องๆช่วยคุณอาเรียพิสูจน์หน่อยว่าคุณอาเรียไปเองหรือมันเป็นเรื่องจริง!

เพราะสำหรับคุณอาเรียแล้ว ทุกคนล้วน เ ท ่ า เ ท ย ม กั น

In [34]: *# Implement code here*

```
sns.set(style="whitegrid")
```

```
# กราฟเปรียบเทียบเงินเดือนตาม Gender
```

```
plt.figure(figsize=(10, 5))
```

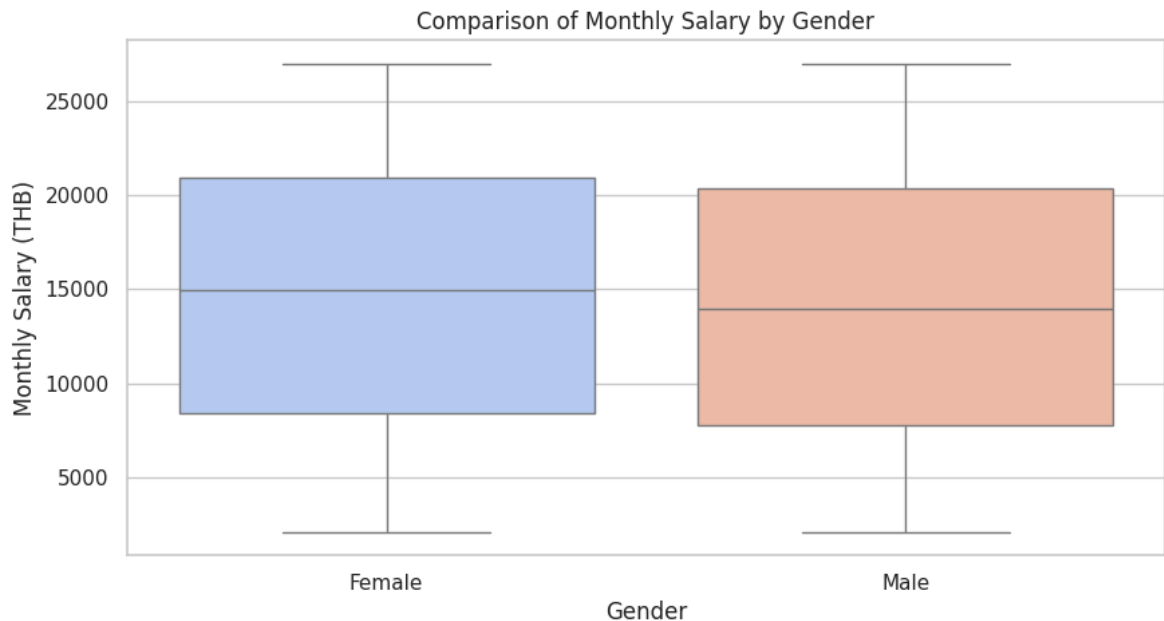
```
sns.boxplot(x="Gender", y="MonthlyRate", data=df, palette="coolwarm")
```

```
plt.title("Comparison of Monthly Salary by Gender")
plt.ylabel("Monthly Salary (THB)")
plt.xlabel("Gender")
plt.show()
```

<ipython-input-34-f65414d84dce>:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v 0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(x="Gender", y="MonthlyRate", data=df, palette="coolwarm")
```



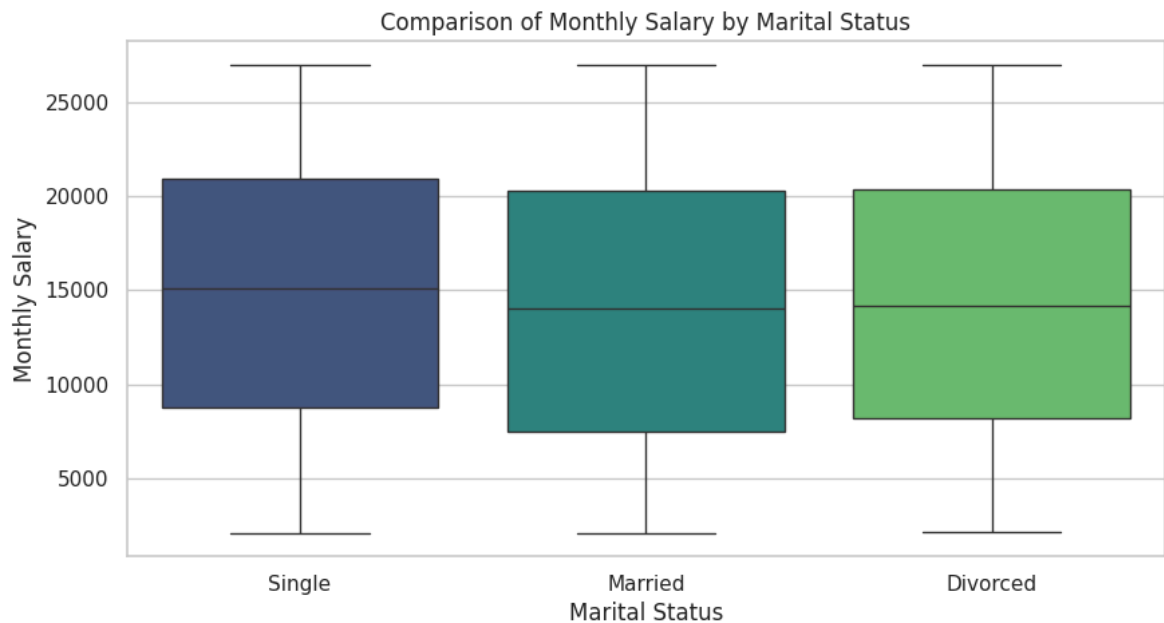
In [42]: *# กราฟเปรียบเทียบเงินเดือนตาม MaritalStatus*

```
plt.figure(figsize=(10, 5))
sns.boxplot(x="MaritalStatus", y="MonthlyRate", data=df, palette="viridis")
plt.title("Comparison of Monthly Salary by Marital Status")
plt.ylabel("Monthly Salary")
plt.xlabel("Marital Status")
plt.show()
```

<ipython-input-42-c8b9e890a31b>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v 0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(x="MaritalStatus", y="MonthlyRate", data=df, palette="viridis")
```

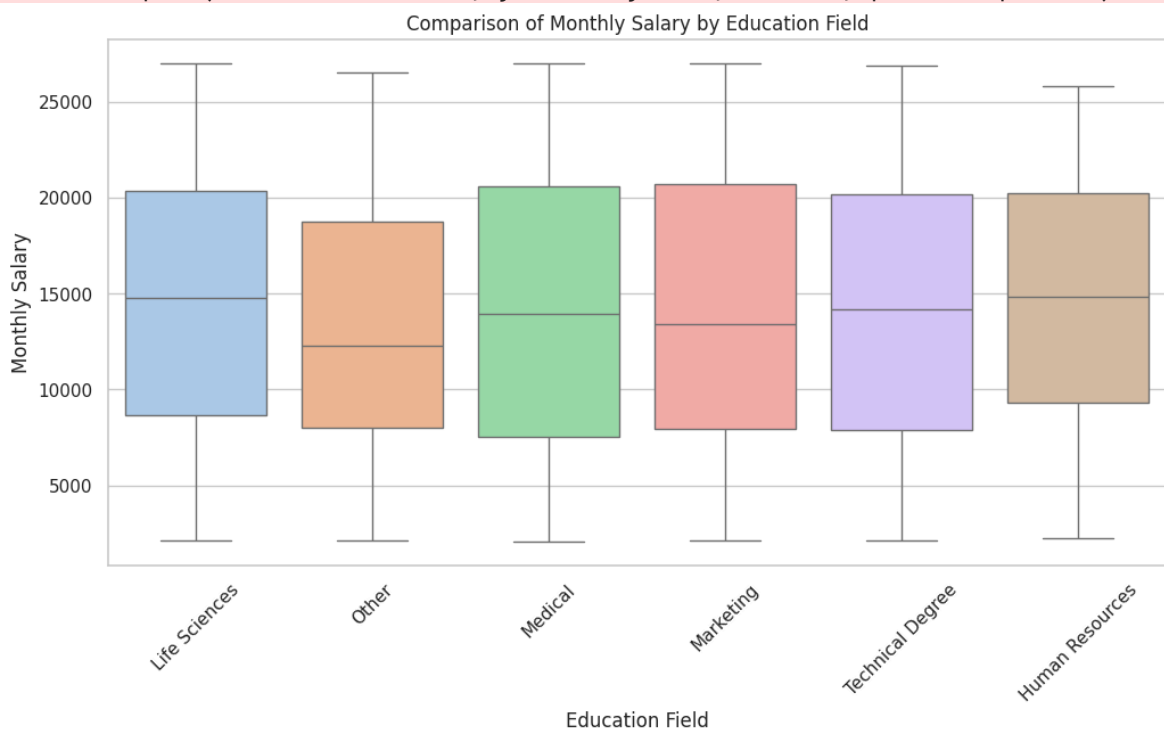


In [33]: `# กราฟเปรียบเทียบเงินเดือนตาม EducationField`
`plt.figure(figsize=(12, 6))`
`sns.boxplot(x="EducationField", y="MonthlyRate", data=df, palette="pastel")`
`plt.xticks(rotation=45)`
`plt.title("Comparison of Monthly Salary by Education Field")`
`plt.ylabel("Monthly Salary")`
`plt.xlabel("Education Field")`
`plt.show()`

<ipython-input-33-1e64b42e9ee3>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v 0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

`sns.boxplot(x="EducationField", y="MonthlyRate", data=df, palette="pastel")`



เนื่องจากโคบายาชิซึ่งจะเป็นคนตรวจโค้ดของน้องๆ

ขอให้น้องๆแสดงข้อมูลอย่างเข้าใจง่ายและแม่นยำ ไม่นะนั้นโคบายาชิซึ่งจะทำการไล่น้องๆออกจากบริษัทนะ ครับ



In [41]: `df.columns`

Out[41]: `Index(['Age', 'DailyRate', 'Department', 'EducationField', 'Gender',
'MaritalStatus', 'MonthlyRate', 'OverTime'],
dtype='object')`

```
In [47]: average_monthly_rate = df.groupby('Gender')['MonthlyRate'].mean().reset_index()

sns.set(style="whitegrid")

plt.figure(figsize=(8, 6))
sns.barplot(x='Gender', y='MonthlyRate', data=average_monthly_rate, palette='pas

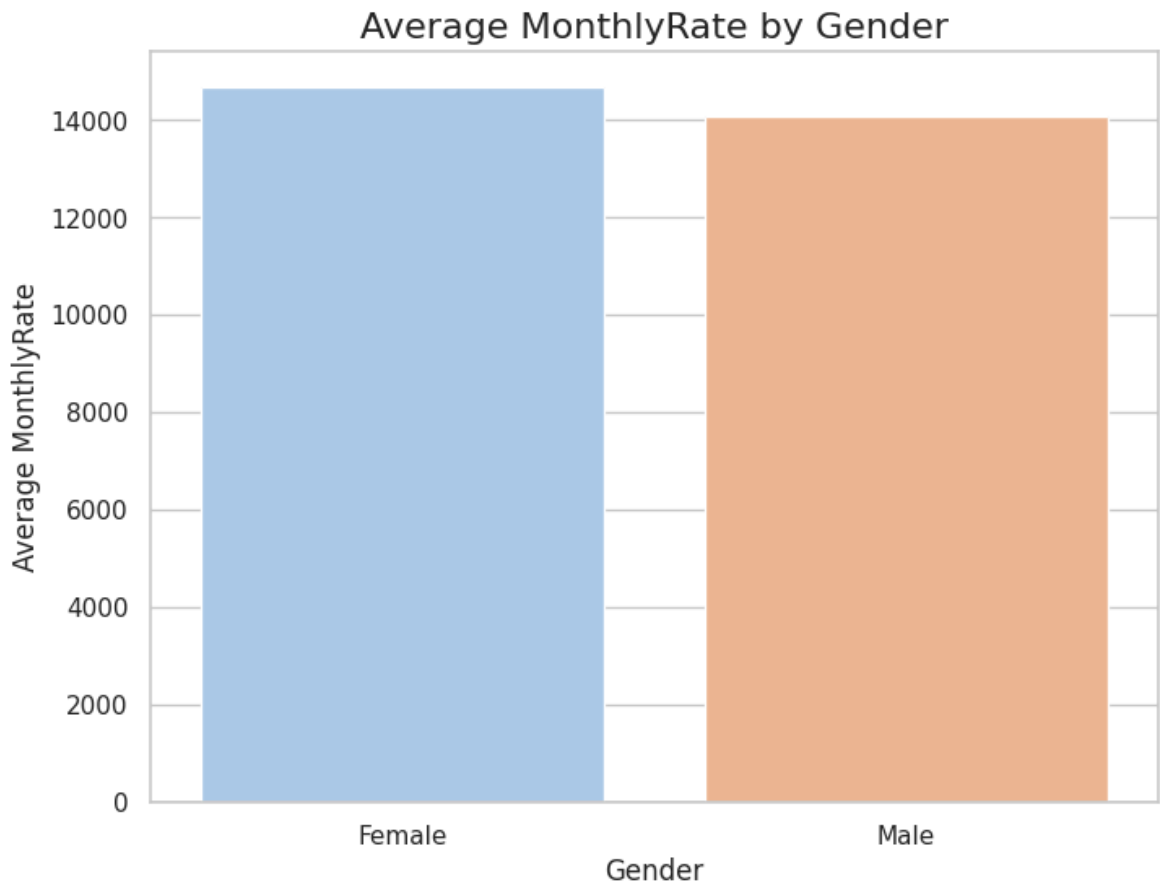
plt.title('Average MonthlyRate by Gender', fontsize=16)
plt.xlabel('Gender', fontsize=12)
plt.ylabel('Average MonthlyRate', fontsize=12)

plt.show()
```

<ipython-input-47-de0fdadb009a>:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v 0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='Gender', y='MonthlyRate', data=average_monthly_rate, palette='pa
stel')
```



```
In [49]: average_monthly_rate = df.groupby('MaritalStatus')['MonthlyRate'].mean().reset_index()

sns.set(style="whitegrid")

plt.figure(figsize=(8, 6))
sns.barplot(x='MaritalStatus', y='MonthlyRate', data=average_monthly_rate, palette='pastel')

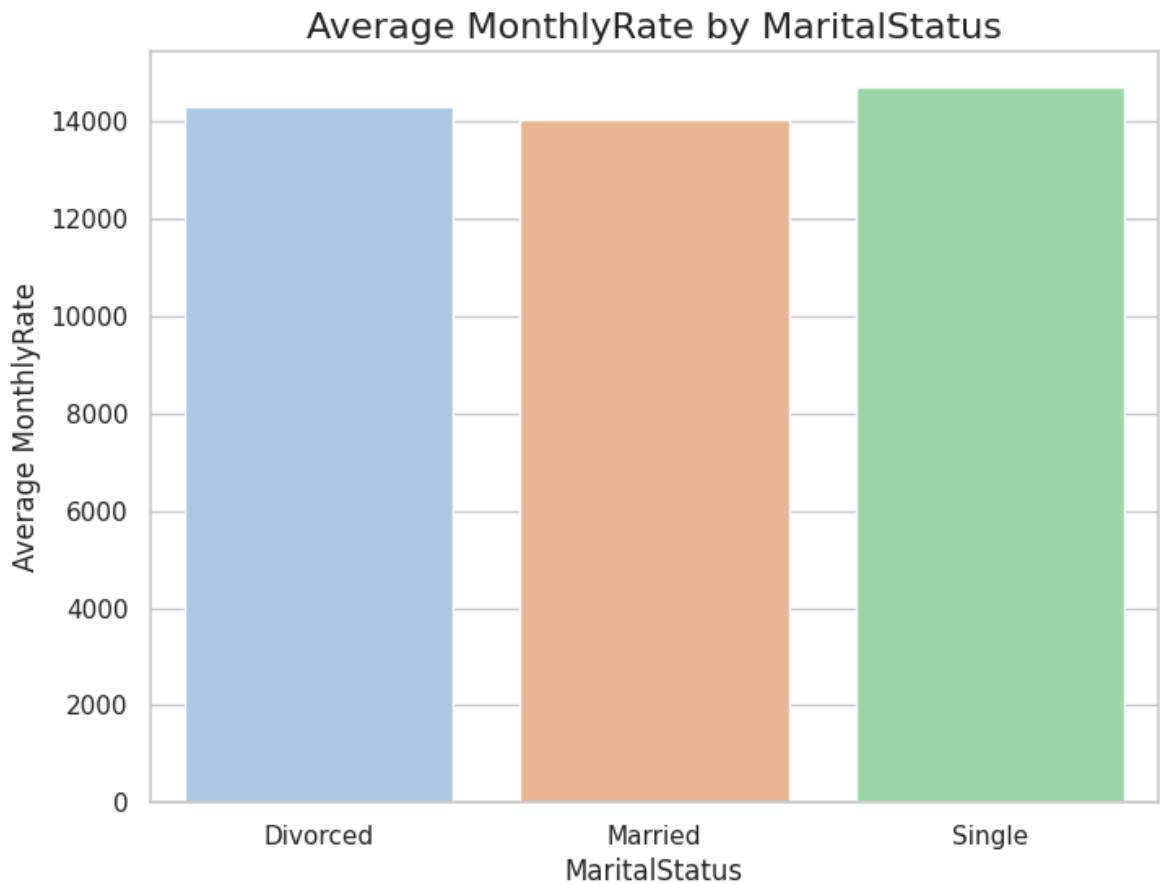
plt.title('Average MonthlyRate by MaritalStatus', fontsize=16)
plt.xlabel('MaritalStatus', fontsize=12)
plt.ylabel('Average MonthlyRate', fontsize=12)

plt.show()
```

<ipython-input-49-cbde7d6bb4c7>:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='MaritalStatus', y='MonthlyRate', data=average_monthly_rate, palette='pastel')
```



```
In [53]: average_monthly_rate = df.groupby('EducationField')['MonthlyRate'].mean().reset_index()

sns.set(style="whitegrid")

plt.figure(figsize=(8, 6))
sns.barplot(x='EducationField', y='MonthlyRate', data=average_monthly_rate, palette='pastel')

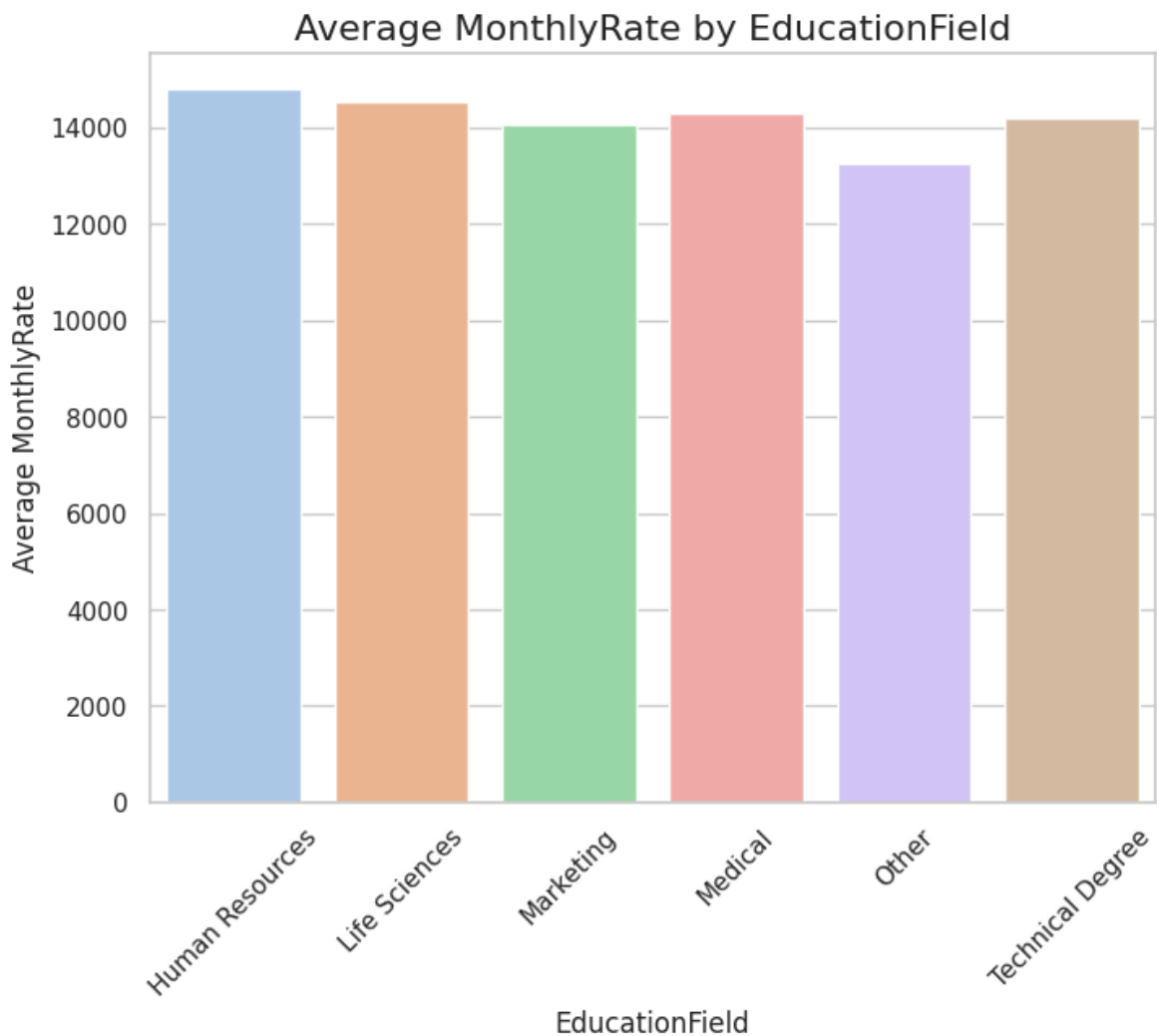
plt.xticks(rotation=45)
plt.title('Average MonthlyRate by EducationField', fontsize=16)
plt.xlabel('EducationField', fontsize=12)
plt.ylabel('Average MonthlyRate', fontsize=12)

plt.show()
```

<ipython-input-53-1adeffd10dda>:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='EducationField', y='MonthlyRate', data=average_monthly_rate, palette='pastel')
```



สรุป

การเปรียบเทียบเงินเดือนตามเพศและสถานะ พบว่าได้รับเงินเดือนที่ใกล้เคียงกัน

การเปรียบเทียบเงินเดือนตามคณะที่เรียน พบว่าได้รับเงินเดือนที่ใกล้เคียงกัน แต่สามารถจำแนกความต่างได้
อยู่ ถ้าดูจากคณะ Human Resources กับ Other

Punchaya Chancharoen
65070507236