



```
In [ ]: # Прочитайте файл 2017_jun_final.csv за допомогою методу read_csv
import pandas as pd
from google.colab import files

# URL для завантаження CSV з Google Sheets
url = 'https://docs.google.com/spreadsheets/d/1bc1X5PbqE8zR1-w0SM3ZTFa53sIC7RXqwtY7Abmw9yQ/e

# Прочитати файл CSV
df = pd.read_csv(url)
```

```
In [ ]: # Прочитайте отриману таблицю, використовуючи метод head
print(df.head())
```

```

      N      Посада Мова програмування Спеціалізація \
0  1      Junior QA engineer      NaN      Manual QA
1  2      Junior QA engineer      NaN      Manual QA
2  3      Software Engineer      C#/.NET      NaN
3  4      Software Engineer      JavaScript      NaN
4  5  Senior Software Engineer      JavaScript      NaN

      Загальний досвід роботи Досвід роботи на поточному місці \
0      0.5      0.5
1      менше 3 місяців      менше 3 місяців
2      4      1
3      5      4
4      10 і більше років      0.5

      Зарплата на місяць Зміна зарплати за 12 місяців Місто Розмір компанії \
0      400      0 Харків до 10 осіб
1      430      130 Харків більше 1000 осіб
2      2500      800 Київ до 1000 осіб
3      2000      400 Одеса до 200 осіб
4      4500      500 Інше до 10 осіб

      ... Ще студент      Рівень англійської \
0  ...      True      Pre-intermediate (нище середнього)
1  ...      True      Upper-Intermediate (вище середнього)
2  ...      False      Intermediate (середній)
3  ...      False      Pre-intermediate (нище середнього)
4  ...      False      Intermediate (середній)

      Предметна область      Дата заповнення \
0      iOS,eCommerce 12/06/2017 12:38:08
1      Веб-розробка 12/06/2017 12:42:51
2  Веб-розробка,Корпоративні системи 12/06/2017 13:02:47
3  Desktop Applications,Веб-розробка 12/06/2017 13:03:01
4  Desktop Applications,Веб-розробка 12/06/2017 13:03:47

      User.Agent      exp current_job_exp \
0  Mozilla/5.0 (Windows NT 10.0; Win64; x64) Appl... 0.5 0.5
1  Mozilla/5.0 (iPad; CPU OS 10_2_1 like Mac OS X... 0.0 0.0
2  Mozilla/5.0 (Windows NT 10.0; Win64; x64) Appl... 4.0 1.0
3  Mozilla/5.0 (Windows NT 6.1; Win64; x64) Apple... 5.0 4.0
4  Mozilla/5.0 (iPhone; CPU iPhone OS 10_3_2 like... 10.0 0.5

      salary Валюта cls
0      400      NaN QA
1      430      NaN QA
2      2500      NaN DEV
3      2000      NaN DEV
4      4500      NaN DEV
```

[5 rows x 25 columns]

```
In [ ]: # Визначте розмір таблиці за допомогою методу shape
print(df.shape)
```

(8704, 25)

```
In [ ]: # Визначте типи всіх стовпців за допомогою dataframe.dtypes
print(df.dtypes)
```

```
N                int64
Посада           object
Мова програмування object
Спеціалізація    object
Загальний досвід роботи object
Досвід роботи на поточному місці object
Зарплата на місяць int64
Зміна зарплати за 12 місяців int64
Місто            object
Розмір компанії object
Тип компанії    object
Стать           object
Вік             int64
Освіта          object
Університет     object
Ще студент      bool
Рівень англійської object
Предметна область object
Дата заповнення object
User.Agent      object
exp            float64
current_job_exp float64
salary          int64
Валюта          object
cls            object
dtype: object
```

```
In [ ]: # Порахуйте, яка частка пропусків міститься в кожній колонці (використовуйте методи isnull та
missing_percentage = df.isnull().sum() / len(df) * 100
print(missing_percentage)
```

```
N                0.000000
Посада           0.000000
Мова програмування 44.232537
Спеціалізація    82.938879
Загальний досвід роботи 0.000000
Досвід роботи на поточному місці 0.000000
Зарплата на місяць 0.000000
Зміна зарплати за 12 місяців 0.000000
Місто            0.000000
Розмір компанії  0.000000
Тип компанії     0.000000
Стать            0.000000
Вік             0.000000
Освіта          0.000000
Університет     4.113051
Ще студент      0.000000
Рівень англійської 0.000000
Предметна область 2.562040
Дата заповнення 0.000000
User.Agent      0.000000
exp            0.000000
current_job_exp 0.000000
salary          0.000000
Валюта          99.885110
cls            18.083640
dtype: float64
```

```
In [ ]: # Видаліть усі стовпці з пропусками, крім стовпця "Мова програмування"
columns_to_keep = ['Мова програмування']
df_no_missing = df.loc[:, df.columns.isin(columns_to_keep) | df.notna().all()]
```

```
In [ ]: # Знову порахуйте, яка частка пропусків міститься в кожній колонці і переконайтеся, що залиш
missing_percentage = df_no_missing.isnull().sum() / len(df_no_missing) * 100
print(missing_percentage)
```

```
print(missing_percentage)
```

```
N                                0.000000
Посада                           0.000000
Мова програмування              44.232537
Загальний досвід роботи         0.000000
Досвід роботи на поточному місці 0.000000
Зарплата на місяць             0.000000
Зміна зарплати за 12 місяців    0.000000
Місто                           0.000000
Розмір компанії                0.000000
Тип компанії                   0.000000
Стать                           0.000000
Вік                             0.000000
Освіта                         0.000000
Ще студент                     0.000000
Рівень англійської            0.000000
Дата заповнення               0.000000
User.Agent                    0.000000
exp                            0.000000
current_job_exp                0.000000
salary                        0.000000
dtype: float64
```

```
In [ ]: # Видалити усі рядки у вихідній таблиці за допомогою методу dropna
df_cleaned = df_no_missing.dropna()
```

```
In [ ]: # Визначити новий розмір таблиці за допомогою методу shape
print(df_cleaned.shape)
```

```
(4854, 20)
```

```
In [ ]: # Створити нову таблицю python_data, в якій будуть тільки рядки зі спеціалістами, які вказал
python_data = df_cleaned[df_cleaned['Мова програмування'] == 'Python']
```

```
In [ ]: # Визначити розмір таблиці python_data за допомогою методу shape
print(python_data.shape)
```

```
(207, 20)
```

```
In [ ]: # Використовуючи метод groupby, виконати групування за стовпчиком "Посада"
grouped = python_data.groupby('Посада')
```

```
In [ ]: # Створіть новий DataFrame, де для згрупованих даних за стовпчиком "Посада", виконайте агрег
salary_stats = grouped['Зарплата на місяць'].agg(['min', 'max']).reset_index()
print(salary_stats)
```

```
      Посада  min  max
0  Junior Software Engineer   100  1250
1  Senior Software Engineer  1300  9200
2    Software Engineer       500  5400
3  System Architect       3000  5000
4  Technical Lead          1425  6200
```

```
In [ ]: # Створіть функцію fill_avg_salary, яка повертатиме середнє значення заробітної плати на міс
def fill_avg_salary(row):
    return (row['min'] + row['max']) / 2
salary_stats['avg'] = salary_stats.apply(fill_avg_salary, axis=1)
```

```
In [ ]: # Створіть описову статистику за допомогою методу describe для нового стовпчика.
print(salary_stats['avg'].describe())
```

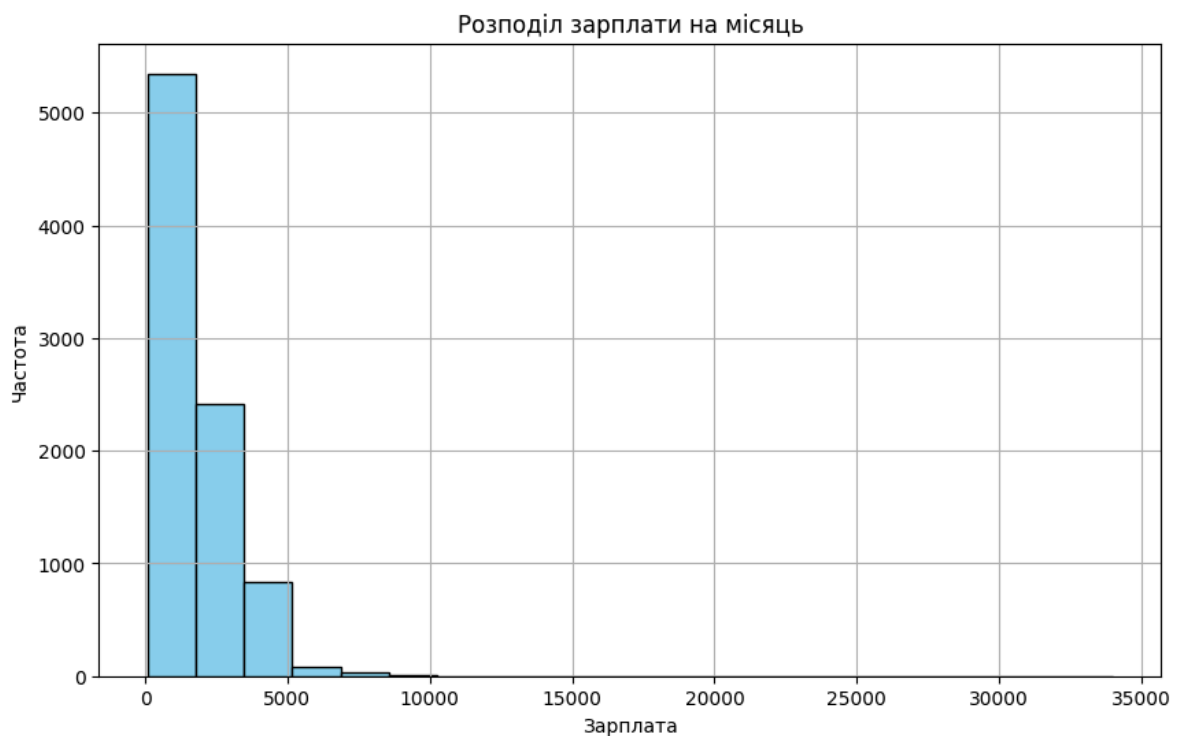
```
count      5.000000
mean     3337.500000
```

```
std      1700.091909
min       675.000000
25%      2950.000000
50%      3812.500000
75%      4000.000000
max      5250.000000
Name: avg, dtype: float64
```

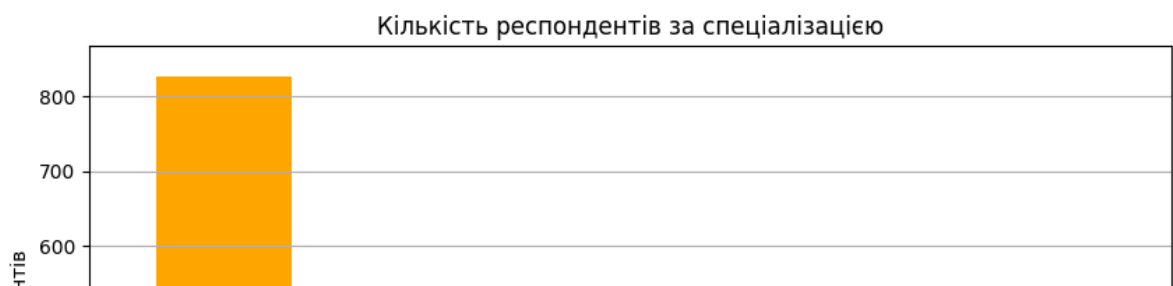
```
In [ ]: # Збережіть отриману таблицю в CSV файл
salary_stats.to_csv('python_salary_stats.csv', index=False)

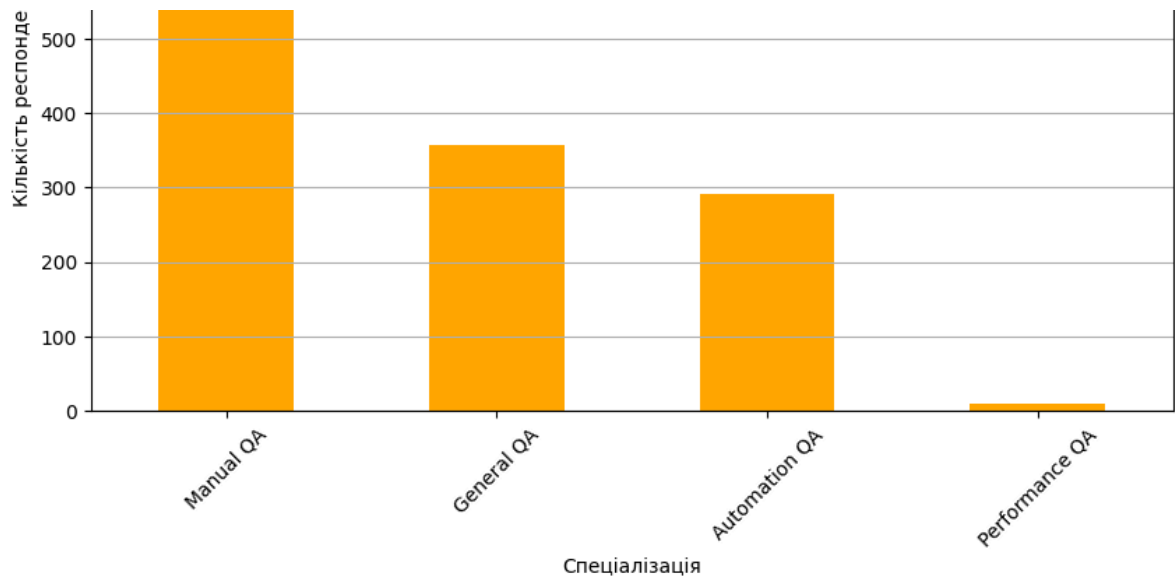
# Завантажити файл на локальний комп'ютер
files.download('python_salary_stats.csv')
```

```
In [ ]: # Гістограма розподілу зарплати на місяць
plt.figure(figsize=(10, 6))
plt.hist(df['Зарплата на місяць'], bins=20, color='skyblue', edgecolor='black')
plt.title('Розподіл зарплати на місяць')
plt.xlabel('Зарплата')
plt.ylabel('Частота')
plt.grid(True)
plt.show()
```



```
In [ ]: # Графік кількості респондентів за кожною спеціалізацією
plt.figure(figsize=(10, 6))
df['Спеціалізація'].value_counts().plot(kind='bar', color='orange')
plt.title('Кількість респондентів за спеціалізацією')
plt.xlabel('Спеціалізація')
plt.ylabel('Кількість респондентів')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.show()
```





```
In [ ]: # Кругова діаграма для розподілу розміру компанії
plt.figure(figsize=(8, 8))
df['Розмір компанії'].value_counts().plot(kind='pie', autopct='%1.1f%%', startangle=140)
plt.title('Розподіл розміру компанії')
plt.axis('equal')
plt.tight_layout()
plt.show()
```

Розподіл розміру компанії

