

Зертханалық жұмыс 7

Визуализация түрлері. Python көмегімен визуализация жасау

Жұмыстың мақсаты: Деректерді визуализациялаудың негізгі дағдыларын игеру. Деректерді визуализациялау-бұл ақпараттың графикалық көрінісі. Деректерді визуализациялау күрделі деректер жиынтығын адам миы оңай түсінетін визуалды форматтарға аударады. Деректерді визуализациялаудың негізгі мақсаты-деректерді қол жетімді және оңай түсіндіруге мүмкіндік беру.

Деректерді визуализациялау үшін қолданылатын құрал:

- Google Collab: <https://colab.research.google.com/>

Мысал ретінде зертханалық жұмыста қолдануға арналған мәліметтер жиынтығы:

- Деректер жиынтығы – population.csv, бұл жиынтық деректерді визуализациялау үшін Google Collab-қа жүктеледі.

Жұмысқа арналған барлық деректер жиынтығы курсқа жүктеледі!

Мысал ретінде жұмыс істеуге арналған негізгі файл: **population.csv**

Жақсы визуализацияны қалай жасауға болады?

- Нақты деректер түрін дұрыс көрсететін диаграмманы таңдаңыз;
- Түстерді пайдаланушыға негізгі көрсеткіштерді тезірек бағалауға көмектесетін етіп пайдаланыңыз;
- Пішіндерге, өлшемдерге, түстердің қанықтылығына назар аударыңыз, бұл ең маңызды нәрсені көрсетеді;
- Деректердегі байланысты түсінуге кедергі келтірмейтін немесе оны бұрмаламайтын визуализация дизайнын таңдаңыз.

Ол үшін :

1. Кітапхананы импорттау керек. Көрнекілік үшін бізге бірнеше кітапхана қажет болады. Осы модульде қолданылатын Python-ның деректерді визуализациялау кітапханалары (Matplotlib, Seaborn) :

- import pandas as pd
- import matplotlib.pyplot as plt

- import seaborn as sns

- import numpy as np

2. csv - df = pd.read_csv('population.csv', thousands=r',') файлын оқу

Google Collab көмегімен кітапханаларды жүктеп алып, файлды оқып шығыңыз, ол үшін Google Collab – қа кіріңіз - Жаңа файлды атаңыз, мазмұнға басыңыз , содан кейін population.csv файлын жүктеу үшін буманы басыңыз. Файл жүктелген кезде, енді белгі бар жерде біз кодты жазамыз, кітапхананы жүктеп және деректерді оқу үшін. Жұмыс істеу үшін, барлық кодты зертханалық жұмыстан көшіруге болады. Кодты жазғыңыз келген сайын мәзірдегі КОД сөзін басыңыз.

The screenshot shows a Google Colab environment. The left sidebar displays the file explorer with 'sample_data' and 'population.csv'. The main area contains a Jupyter Notebook with the following code:

```
[1] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

df = pd.read_csv('population.csv', thousands=r',')

[3] df
```

The output of the code is a preview of the 'population.csv' file, showing a table with 10 columns: Районы, Общее население, мужчины, женщины, городское население, городское население мужчины, городское население женщины, сельское население, сельское население мужчины, and сельское население женщины. The table contains data for various districts, including Semei c.a., Kurchatov c.a., Abai district, Aksuat district, Ayagoz district, Abay district, and AlFarabi district.

Визуализация түрлері

Енді деректерді визуализациялаймыз :

1. Жалпы халықтың таралу(орналасу) гистограммасы

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

```
sns.histplot(df["Общее население"], bins=20, kde=True, color='skyblue')
```

```
plt.xlabel("Общее население")
```

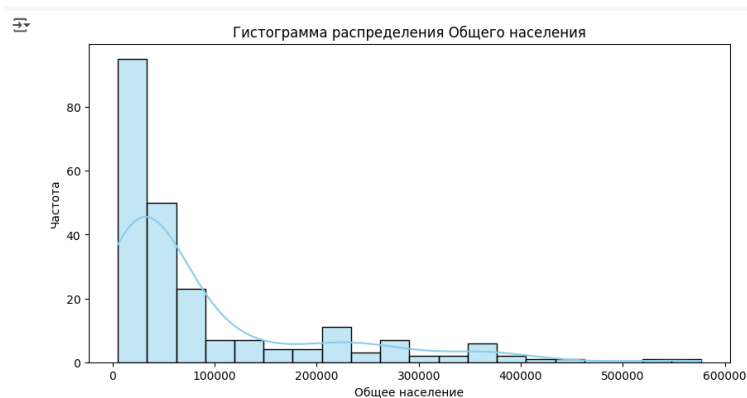
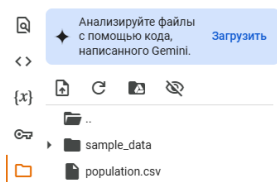
```
plt.ylabel("Частота")
```

```
plt.title("Гистограмма распределения Общего населения")
```

```
plt.show()
```

The screenshot shows the same Google Colab environment as before, but now with the execution of the histogram code. The code is as follows:

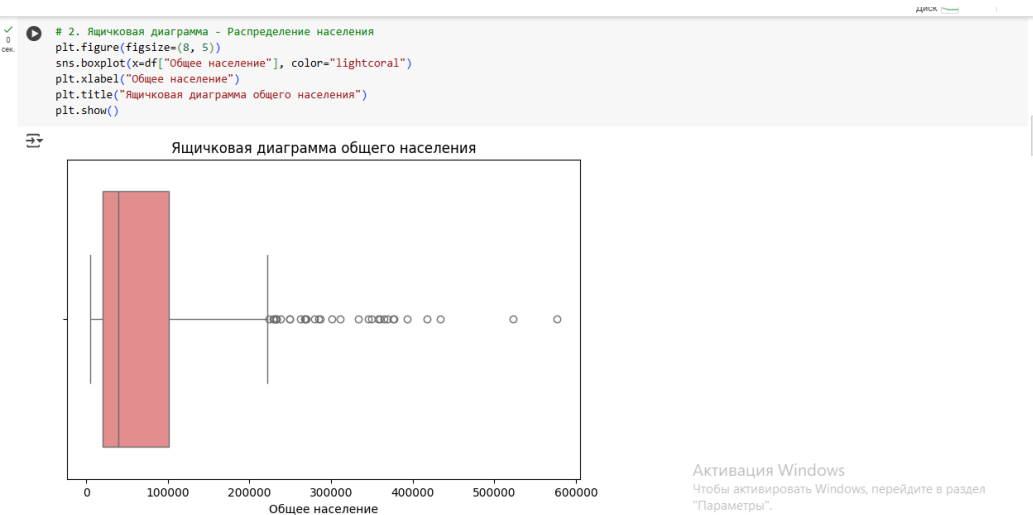
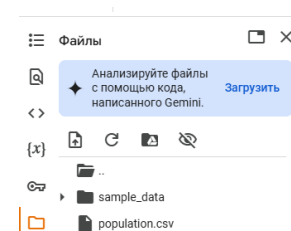
```
# 1. Гистограмма распределения Общего населения
plt.figure(figsize=(10, 5))
sns.histplot(df["Общее население"], bins=20, kde=True, color='skyblue')
plt.xlabel("Общее население")
plt.ylabel("Частота")
plt.title("Гистограмма распределения Общего населения")
plt.show()
```



Активация Windows

2. Қорап диаграммасы-Халықтың таралуы бойынша

```
plt.figure(figsize=(8, 5))
sns.boxplot(x=df["Общее население"], color="lightcoral")
plt.xlabel("Общее население")
plt.title("Ящичковая диаграмма общего населения")
plt.show()
```

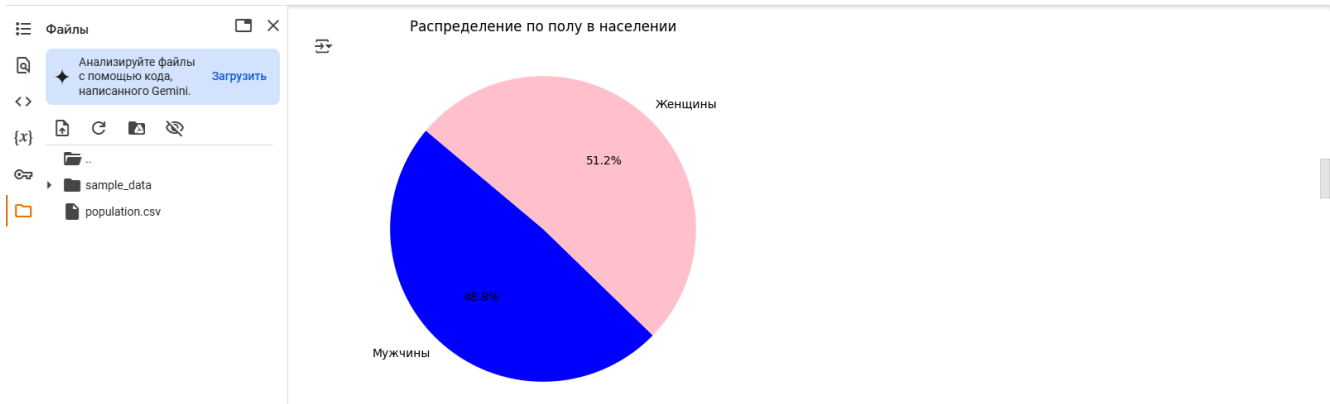


Активация Windows
Чтобы активировать Windows, перейдите в раздел "Параметры".

✓ 14 сек. выполнено в 19:19

3. Дөңгелек диаграмма-Жынысы бойынша

```
gender_counts = [df["мужчины"].sum(), df["женщины"].sum()]
labels = ["Мужчины", "Женщины"]
colors = ["blue", "pink"]
plt.figure(figsize=(6, 6))
plt.pie(gender_counts, labels=labels, autopct="%1.1f%%", colors=colors,
startangle=140)
plt.title("Распределение по полу в населении")
plt.show()
```



4. Бағаналы диаграмма-қала мен ауыл тұрғындарын салыстыру

```
urban_population = df["городское население"].sum()
```

```
rural_population = df["сельское население"].sum()
```

```
plt.figure(figsize=(8, 5))
```

```
sns.barplot(x=["Городское население", "Сельское население"],
```

```
y=[urban_population, rural_population], palette=["green", "brown"])
```

```
plt.ylabel("Общее население")
```

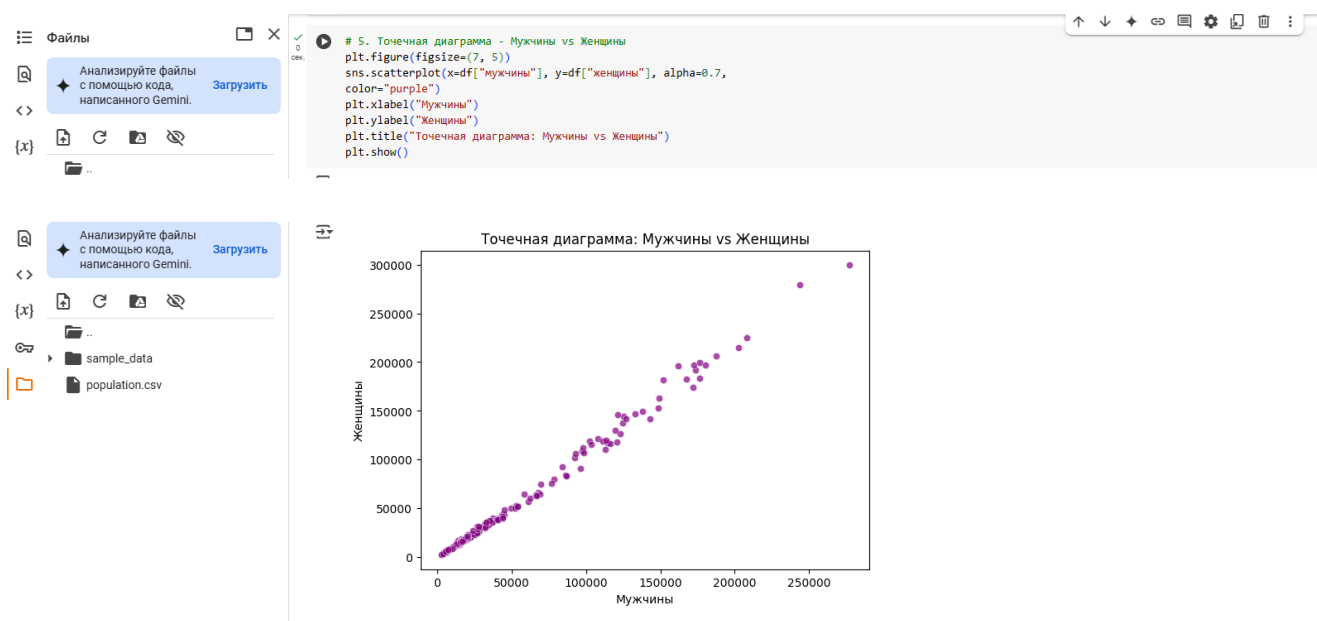
```
plt.title("Сравнение городского и сельского населения")
```

```
plt.show()
```



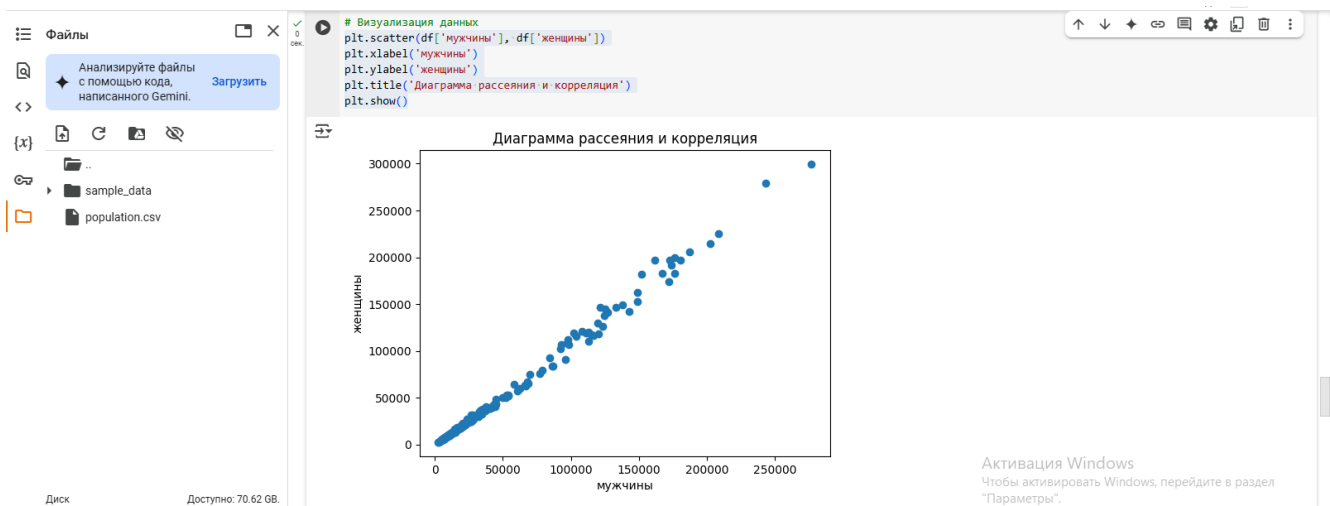
5. Нүктелік диаграмма-ерлер мен әйелдер бойынша

```
plt.figure(figsize=(7, 5))
sns.scatterplot(x=df["мужчины"], y=df["женщины"], alpha=0.7,
color="purple")
plt.xlabel("Мужчины")
plt.ylabel("Женщины")
plt.title("Точечная диаграмма: Мужчины vs Женщины")
plt.show()
```

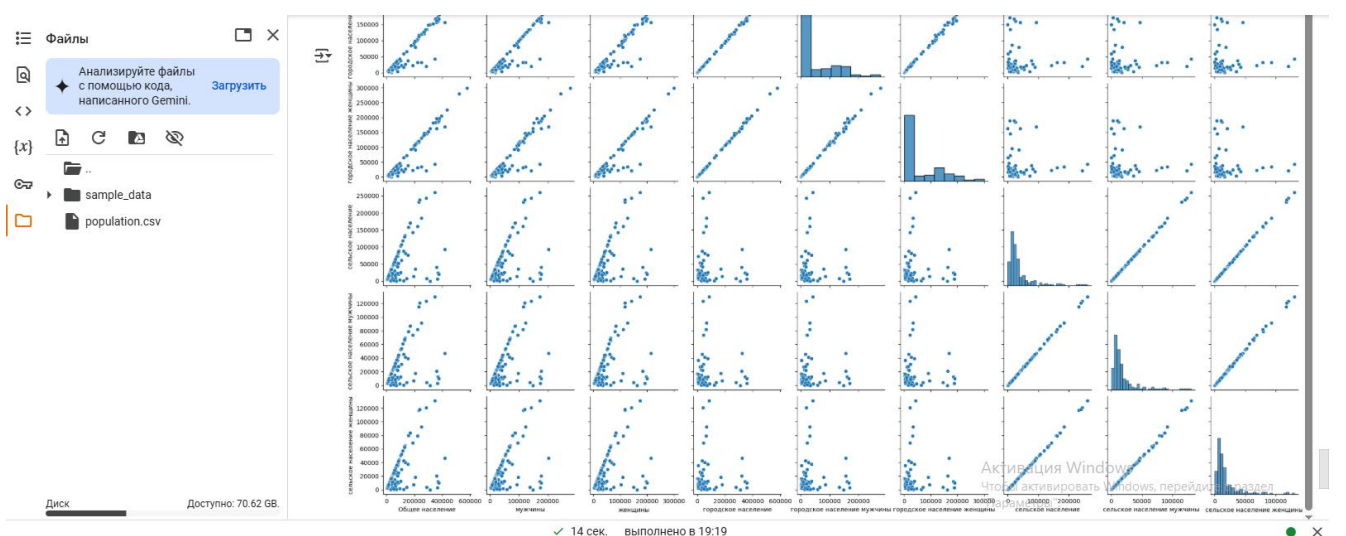
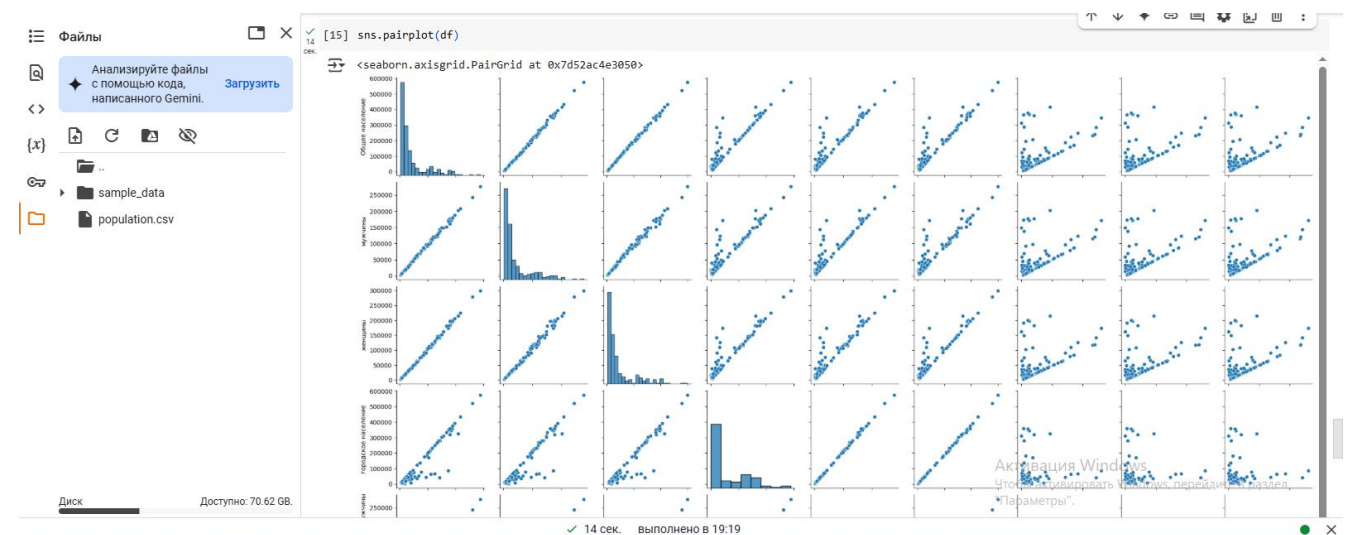


6. Деректерді визуализациялау. Корреляциялық талдау

```
plt.scatter(df['мужчины'], df['женщины'])
plt.xlabel('мужчины')
plt.ylabel('женщины')
plt.title('Диаграмма рассеяния и корреляция')
plt.show()
```

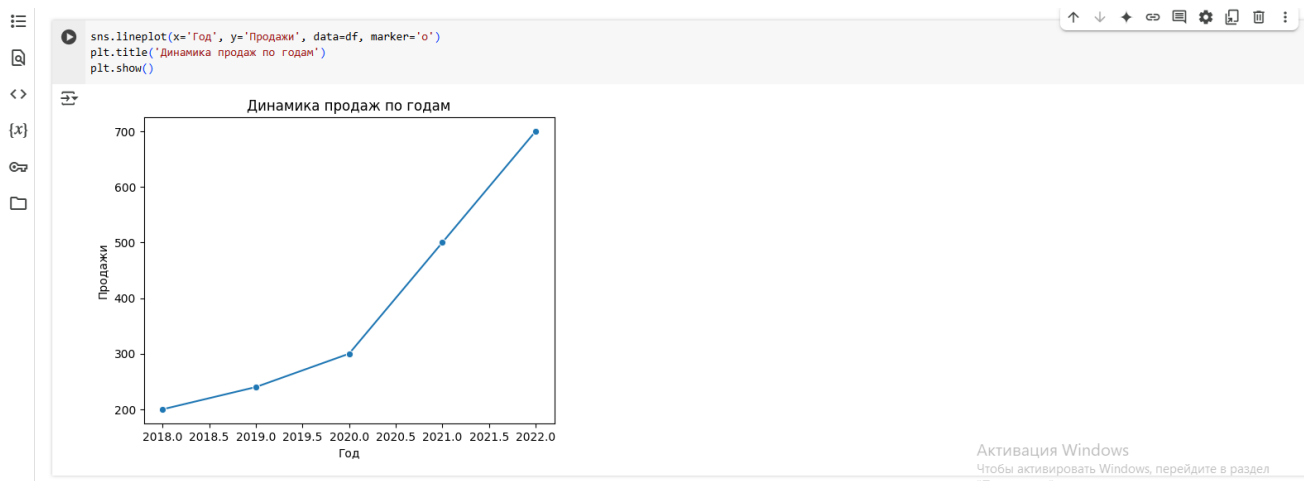


7. Деректерді визуализациялау. Барлық айнымалылардың тәуелділіктерін визуализациялау `sns.pairplot(df)` – осы кодты қолдана отырып, сіз бір визуализациядағы барлық көрсеткіштерге тәуелділіктер аласыз.

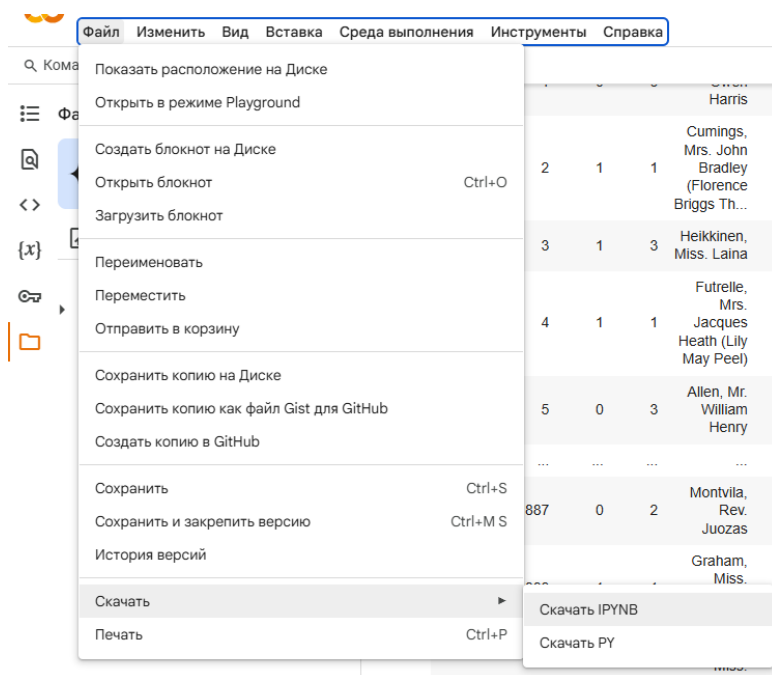


8. Сызықтық график. Мысал үшін сызықтық график басқа мәліметтермен салынған.

```
sns.lineplot(x='Год', y='Продажи', data=df, marker='o')
plt.title('Динамика продаж по годам')
plt.show()
```



Соңында, барлық визуализациялар аяқталғаннан кейін файлды сақтау керек. Ол үшін мәзірден таңдаңыз Файл-жүктеу-жүктеу IPYNB(Файл-Скачать-Скачать IPYNB). Файлға атау беріңіз және сақтаңыз. Дәл осы файлды тексеруге жіберу керек.



P.S.: Назар аударыңыз, мысалы, сіз Google Collab-та жұмыс істеп, содан кейін жұмысты тоқтатып, келесі күні кодты ашу арқылы жалғастыруды шешкен кезде (барлық код Google Drive-та Google Collab бұмасында автоматты түрде сақталады), алдымен population.csv, файлын қайтадан жүктеп алуыңыз керек немесе сіз жұмыс істейтін файл атауы өзгеруі мүмкін. Жүктелгеннен кейін Мәзірден таңдау арқылы- Барлық кодты қайтадан іске қосу(Меню-Среда выполнения-Выполнить все) керек-жұмыс уақыты-бәрін орындаңыз немесе оны басу арқылы барлық кодты қолмен іске қосып, белгілерді басу арқылы. Өзіңізге қаншалықты ыңғайлы солай таңдау

арқылы.

Жеке практикалық тапсырма.

Мақсаты: Деректерді талдау үшін Matplotlib және Seaborn кітапханаларын пайдаланып графиктердің әртүрлі түрлерін құруды үйрену.

ТАПСЫРМА

Titanic деректер жинағын (titanic.csv файлын) жүктеу үшін pandas кітапханасын пайдаланыңыз. Бұл сіз бұрын зертханалық 6-да жұмыс істеген таныс файл. Файл осы тапсырманың астындағы ДОТ платформасына жүктелген.

1. Деректерді дайындау

1.1. Титаник жолаушылар деректерін жүктеп алыңыз:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Деректерді жүктеп алып
df = pd.read_csv("titanic.csv")
df.head()
```

1.2. Деректерді өткізіп алған мәндерді тексеріп, оларды өңдеңіз (мысалы, жою немесе ауыстыруды орындау арқылы).

Сіз алдыңғы жұмысқа қараңыз, өткізіп алған мәндерді алып тастаған және оларды өңдеңізге.

2. Визуализацияны құру

Келесі сұрақтарға жауап беретін визуализация құрыңыз.

2.1. Жолаушылар жасы бойынша гистограмма .

plt.hist() немесе **sns.histplot()** пайдаланып

2.2. Кабина(каюта) кластары бойынша жолаушылардың жасын салыстыруға арналған қорап диаграммасы(boxplot).

sns.boxplot(x='класс', y='Возраст', data=df) пайдаланып

2.3 Тірі қалғандар мен өлгендердің үлесін бейнелеуге арналған дөңгелек

диаграмма.

plt.pie() әдісін пайдаланып

2.4 Жолаушылар санының жасына байланысты сызықтық графигі.

sns.lineplot() пайдаланып

3. Қосымша визуализация бойынша талдау

3.1. Бірнеше сандық бағандар үшін **sns.pairplot()** жұптық диаграммасын жасап көріңіз.

3.2. Кез келген екі айнымалының өзара байланысын талдау үшін тренд сызығы бар **scatterplot** нүктелік диаграммасын жасаңыз.

4. Графиктерді ресімдеуге қойылатын талаптар

4.1.**plt.title(), plt.xlabel(), plt.ylabel()** осьтерінің тақырыбы мен атауын (заголовки и подписи) қосыңыз

Қандай түрде жұмыс өткізіледі : 1. Аяқталған тапсырманы Google Colab-та орындау ұсынылады. Жұмысты аяқтағаннан кейін блокнотты (блокнот .іруnb) жүктеу керек Google Colab көмегімен және оны тапсырыңыз. 2. Сондай-ақ, әр тармақ бойынша тапсырманың орындалуы бойынша түсініктемелер мен жауаптары бар скриншоттармен есеп беру қажет. Скриндар курстағы талаптарға сай болуы керек, жүйелік жұмыс уақыты мен күнімен толық болуы керек. Егер скриндар талаптарға сәйкес келмесе, бағалау 50% - ға төмендейді.

Соңында сіз екі файлды, бір файлды *.docx форматында тіркеуіңіз керек, келесі бір файлды *.іруnb форматта, файлдарды архивке тіркемеңіз, өйткені архив әрдайым ашылмайды.