

Отчёт по практической работе №4

Визуализация данных средствами Matplotlib. Основы

Цель работы:

Получить навыки использования библиотеки визуализации данных Matplotlib с использованием языка Python.

Ход работы:

Согласно указаниям в методичке создадим блокнот на сайте Google Colab

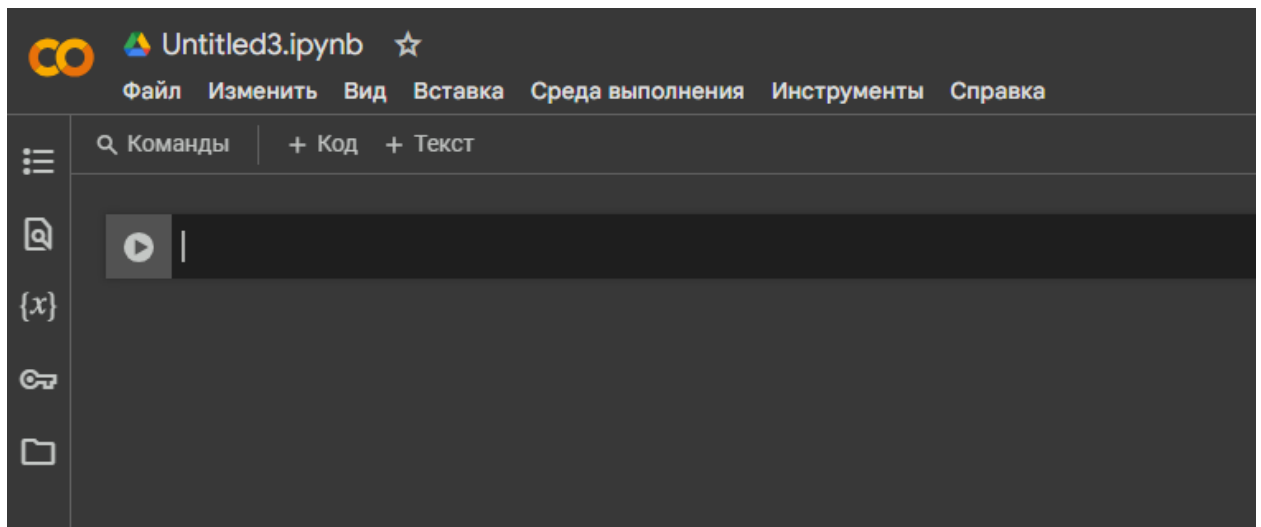


Рис 4.1.

Скопируем код из методички для проверки работы

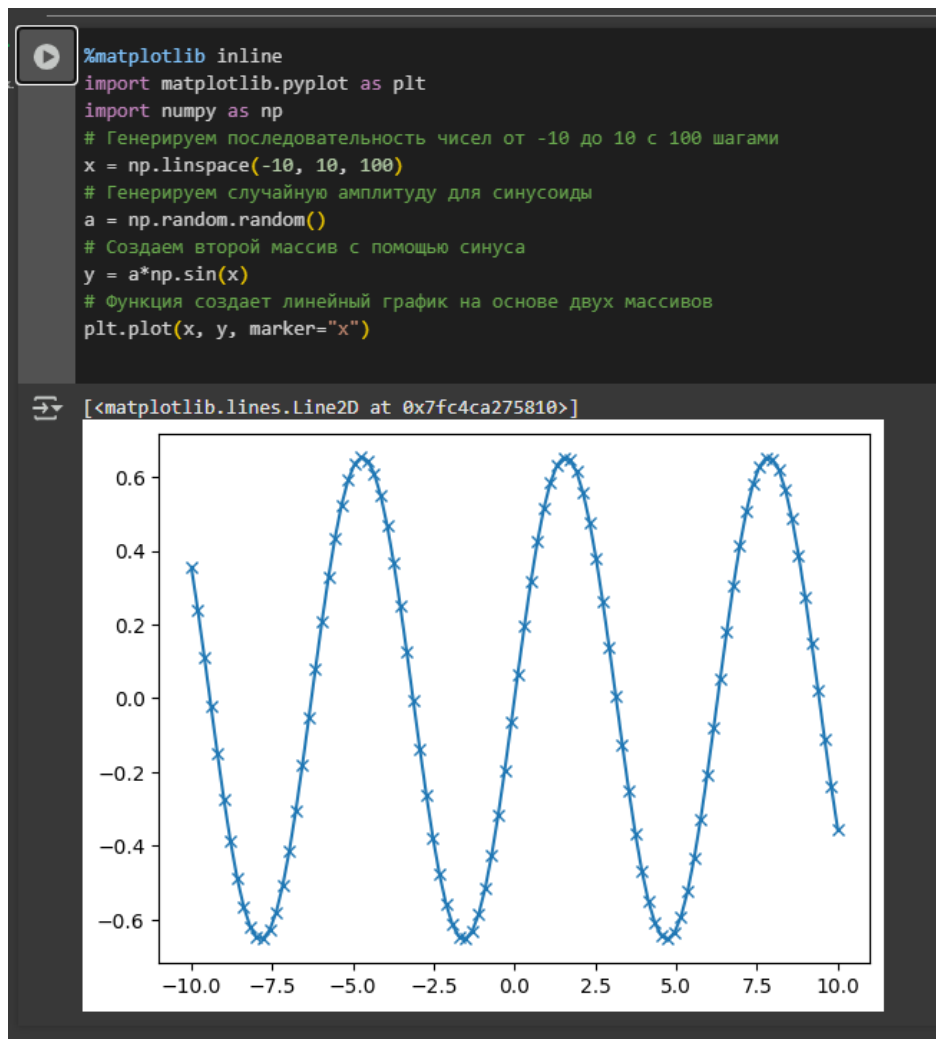


Рис 4.2.

После того как график выведен и отработал правильно перейдём к основной части. После копирования кода он выдал ошибку поэтому проверим код и отформатируем его, добавив комментарии

```

# Загружаем данные из Excel
df_can = pd.read_excel(
    'https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/DV0101EN/labs/Data_Files/Canada.xlsx',
    sheet_name='Canada by Citizenship',
    skiprows=range(20),
    skipfooter=2
)
print('Данные загружены и записаны в dataframe!')

# Выводим первые строки таблицы
df_can.head()

# Выводим размерность таблицы
print(df_can.shape)

# Удаляем лишние столбцы
df_can.drop(['AREA', 'REG', 'DEV', 'Type', 'Coverage'], axis=1, inplace=True)
df_can.head()

# Переименовываем столбцы
df_can.rename(columns={'OdName': 'Country', 'AreaName': 'Continent', 'RegName': 'Region'}, inplace=True)
df_can.head()

# Проверим, все ли столбцы строковые
all(isinstance(column, str) for column in df_can.columns)

# Преобразуем столбцы в строки
df_can.columns = list(map(str, df_can.columns))

# Проверим снова
all(isinstance(column, str) for column in df_can.columns)

# Устанавливаем индекс по стране
df_can.set_index('Country', inplace=True)
df_can.head()

# Добавляем столбец Total с суммой по строкам
# Преобразуем все данные в числовые, где возможно (нечисловые заменятся на NaN)
df_can[years] = df_can[years].apply(pd.to_numeric, errors='coerce')

# Теперь можно добавить столбец Total с суммой по строкам
df_can['Total'] = df_can[years].sum(axis=1)
df_can.head()

df_can.head()

# Создаём список годов для сортировки
years = list(map(str, range(1980, 2014)))

# Сортируем данные по столбцу Total
df_can.sort_values(['Total'], ascending=False, axis=0, inplace=True)
df_top5 = df_can.head()

# Транспонируем таблицу для топ-5 стран
df_top5 = df_top5[years].transpose()
df_top5.head()

# Устанавливаем стиль графика ggplot
mpl.style.use('ggplot')

# Проверим версию Matplotlib
print('Matplotlib version: ', mpl.__version__) # >= 2.0.0

```

Рис 4.3

Код выдавал ошибки поэтому были внесены изменения

```

# Преобразуем все данные в числовые, где возможно (нечисловые заменятся на NaN)
df_can[years] = df_can[years].apply(pd.to_numeric, errors='coerce')

# Теперь можно добавить столбец Total с суммой по строкам
df_can['Total'] = df_can[years].sum(axis=1)
df_can.head()

```

Рис 4.4.

После этих изменений получили вот такой график

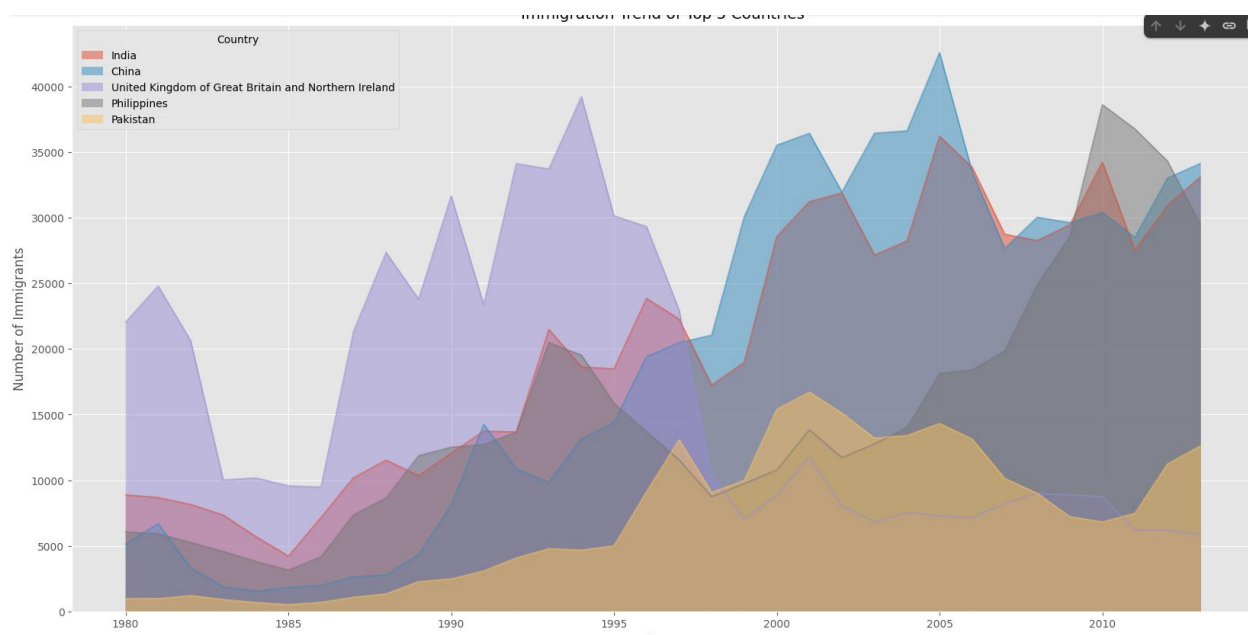


Рис 4.5.

Так же сохраним данную работу так как она понадобится нам в работе №5

Вывод:

В данной работе мы изучили основы построения графиков с помощью Matplotlib. Так же решили проблему с табуляциями и дополнили программу комментариями для улучшения читаемости текста.