**Работа с библиотекой Seaborn**

**Seaborn** — это библиотека для создания статистических графиков на Python. Она основывается на matplotlib и тесно взаимодействует со структурами данных pandas.

Архитектура Seaborn позволяет вам быстро изучить и понять свои данные. Seaborn захватывает целые фреймы данных или массивы, в которых содержатся все ваши данные, и выполняет все внутренние функции, нужные для семантического маппинга и статистической агрегации для преобразования данных в информативные графики.

Она абстрагирует сложность, позволяя вам проектировать графики в соответствии с вашими нуждами.

Установить seaborn можно так же, как и любую другую библиотеку, для этого понадобится менеджер пакетов Python. Во время установки seaborn библиотека установит все зависимости, включая matplotlib, pandas, numpy и scipy.

Установим seaborn и, конечно же, также пакет notebook, чтобы получить доступ к песочнице с данными.

pipenv install seaborn notebook

Помимо этого, перед началом работы импортируем несколько модулей.

import seaborn as sns

import pandas as pd

import numpy as np

import matplotlib

**Строим первые графики**

Seaborn  работает непосредственно с объектами dataframe из pandas, что делает ее очень удобной. Библиотека поставляется с некоторыми встроенными наборами данных, которые можно использовать прямо из кода, и не загружать файлы вручную.

Посмотрим, как это работает на наборе данных о рейсах самолетов.

flights\_data = sns.load\_dataset("flights")

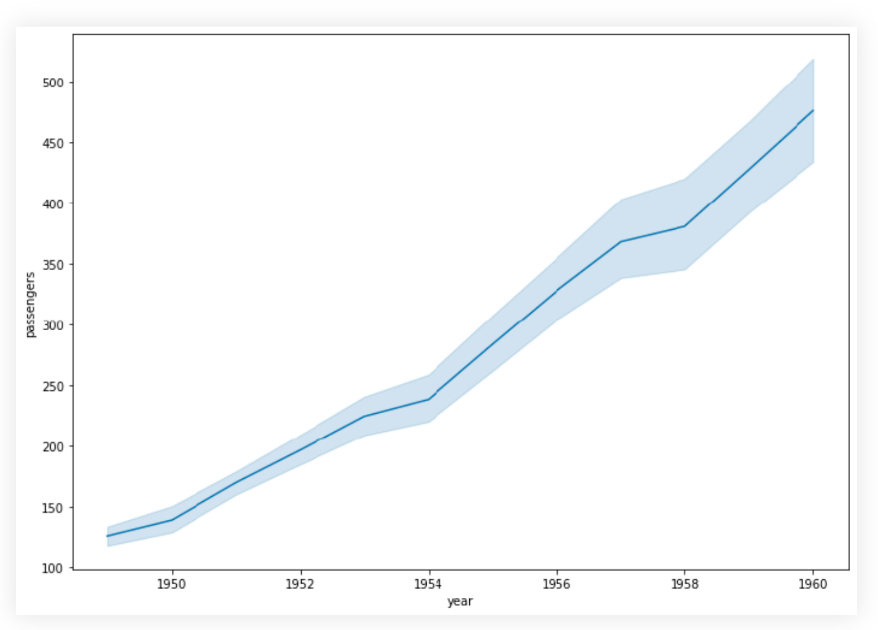
flights\_data.head()

|  |  |  |  |
| --- | --- | --- | --- |
|  | year | month | passengers |
| 0 | 1949 | Jan | 112 |
| 1 | 1949 | Feb | 118 |
| 2 | 1949 | Mar | 132 |
| 3 | 1949 | Apr | 129 |
| 4 | 1949 | May | 121 |

Функция load\_dataset, ожидает имя загружаемых данных и возвращает dataframe. Все эти наборы данных доступны в репозитории на Github.

sns.lineplot(data=flights\_data, x="year", y="passengers")

Этот график рисует линию, которая представляет собой развитие непрерывных или категориальных данных. Этот вид графиков популярен и известен, и его легко создать. Воспользуемся функцией lineplot с набором данных и столбцами, представляющими оси x и y.



**Задания на практическую работу**

**Часть 1**

1. Импортируйте данные из файла prog\_workers.csv, содержащем информацию о работниках, стране проживания, дате рождения, заработной плате, опыте работы и предпочитаемом языке программирования

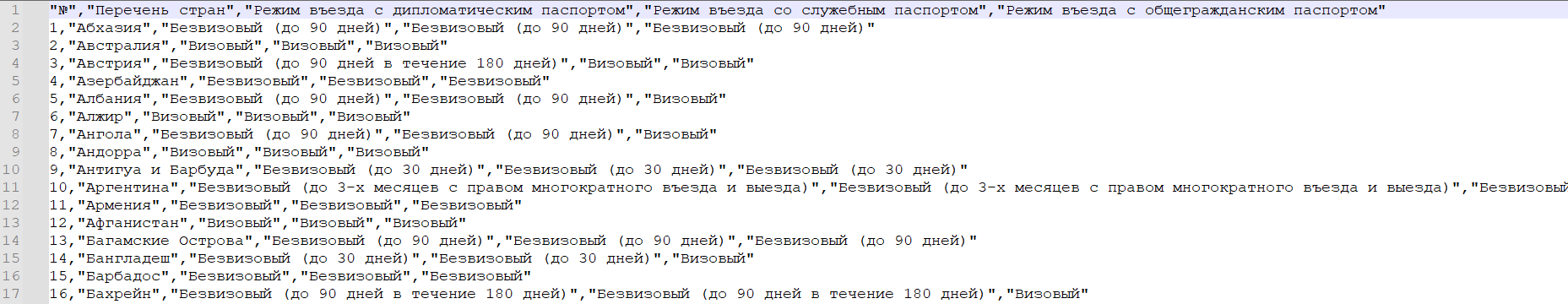
Изображение выглядит как текст

Автоматически созданное описание

1. Постройте гистограмму сотрудников с зарплатой выше 15000 (синим цветом) и ниже 15000 (красным цветом).
2. Постройте гистограмму по количеству работников с разными языками программирования.
3. Постройте гистограмму по средним зарплатам по каждому языку программирования.
4. Постройте тепловую карту для отображения взаимосвязи между языком программирования и опытом работы.
5. Постройте диаграмму рассеяния по возрасту работников и заработной плате.

**Часть 2**

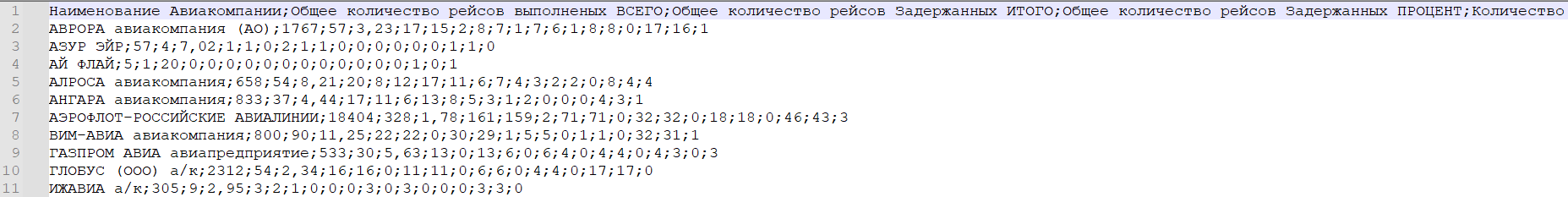
1. Скачайте файл Visa\_countries.csv, содержащий информацию о странах и режимах въезда на их территорию



1. Создайте горизонтальную гистограмму по режимам въезда на территорию стран и количеством этих стран. Столбцы гистограммы должны быть сгруппированы.
2. По данным количества стран и сроков пребывания постройте «скрипичную» диаграмму.

**Часть** 3

1. Импортируйте данные из файла flight delays.csv, содержащего данные о задержках рейсов.



1. Найдите 10 авиакомпаний, сделавших наибольшее количество рейсов всего)
2. Для них постройте гистограмму по числу задержанных рейсов (по группам).
3. Для всех данных постройте объемную диаграмму рассеяния по Общему числу выполненных рейсов, Общему числу задержанных рейсов, Количеству рейсов задержанных более двух часов.
4. Постройте тепловую диаграмму корреляции между Общим числом выполненных рейсов и Количеству рейсов задержанных более шести часов ВСЕГО (для первых 20 записей).