**Московский государственный технический**

**университет им. Н.Э. Баумана.**

Факультет «Информатика и управление»

Кафедра ИУ5. Курс «Машинное обучение»

Отчет по лабораторной работе №1

«Разведочный Анализ данных. Исследование и визуализация данных»

|  |  |  |
| --- | --- | --- |
| Выполнил: |  | Проверил: |
| студент группы РТ5-61Б |  | преподаватель каф. ИУ5 |
| Гудилин Дмитрий |  | Гапанюк Ю. Е. |
|  |  |  |
| Подпись и дата: |  | Подпись и дата: |

Москва, 2020 г.

**Описание задания лабораторной работы**

 Построение основных графиков, входящих в этап разведочного анализа данных.

Для выполнения текущей лабораторной работы мы возьмем dataset <https://www.kaggle.com/luthfim/steam-reviews-dataset>

В нем анализируются данные по отзывам из платформы steam.

**Общий анализ**

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

sns.set(style="ticks")

data = pd.read\_csv(r'C:\Users\dimag\Desktop\Mash\_learning\steam\_reviews.csv', sep=",")

data.head()



data.shape



print('Всего строк: {}'.format(total\_count))



total\_count = data.shape[0]

print('Всего строк: {}'.format(total\_count))



data.columns

Index(['date\_posted', 'funny', 'helpful', 'hour\_played',

'is\_early\_access\_review', 'recommendation', 'review', 'title'],

dtype='object')

data.dtypes

date\_posted object

funny int64

helpful int64

hour\_played int64

is\_early\_access\_review bool

recommendation object

review object

title object

dtype: object

data.isnull().sum()

date\_posted 0

funny 0

helpful 0

hour\_played 0

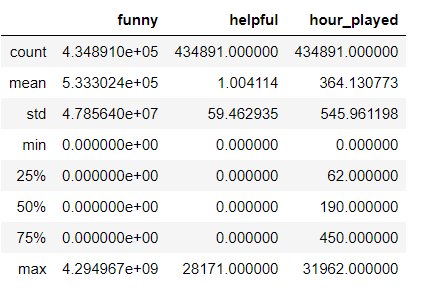
is\_early\_access\_review 0

recommendation 0

review 1516

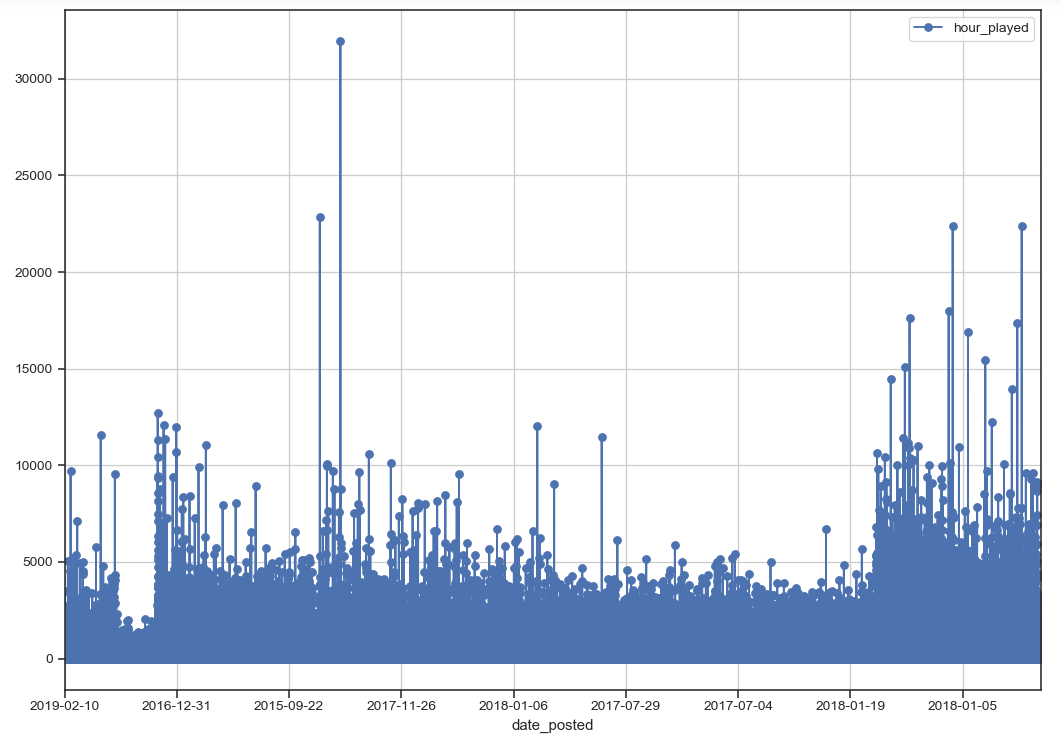
title 0

data.describe()

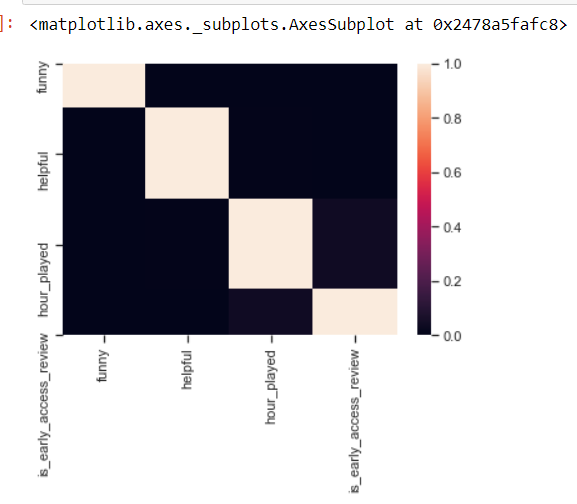


**Графическая часть**

Зависимость наигранных часов от даты поста



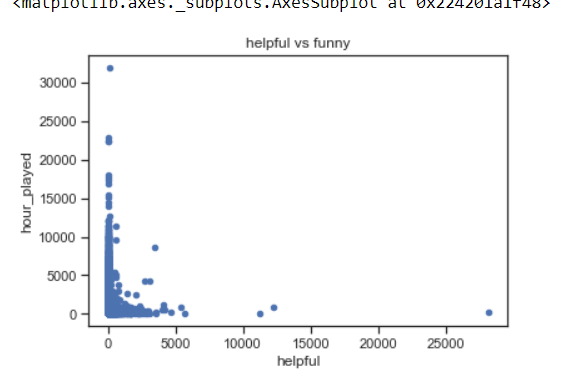
Тепловая карта



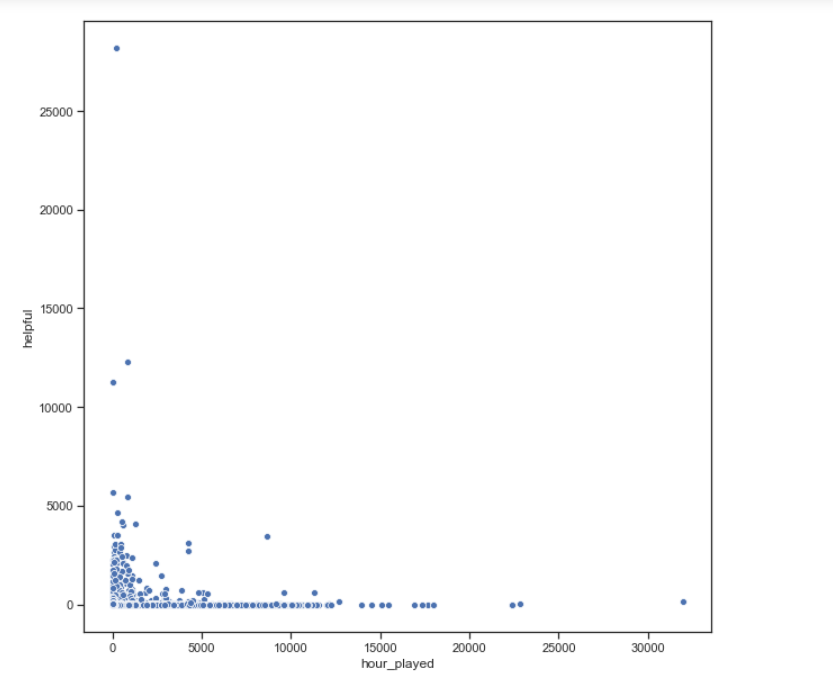
fig, ax = plt.subplots(figsize=(10,10))

sns.scatterplot(ax=ax, x='hour\_played', y='helpful', data=data)

Карта распределения



data.plot(kind='scatter', x='helpful',y='hour\_played',title='helpful vs funny')



Распределение вероятности

