**Московский государственный технический университет**

**им. Н.Э. Баумана**

**(МГТУ им. Н.Э. Баумана)**

**Радиотехнический факультет (РТ)**

Отчёт по лабораторной работе № 1

По теме

«Разведочный анализ данных. Исследование и визуализация данных»

По дисциплине

«Технологии машинного обучения»

|  |  |
| --- | --- |
| Проверил: | Выполнил: |
| Преподаватель кафедры ИУ-5 | студент группы РТ5-61Б |
| Гапанюк Ю.Е. | Лычагин Д.А. |
| Подпись: \_\_\_\_\_\_\_\_\_\_\_\_\_ | Подпись: \_\_\_\_\_\_\_\_\_\_\_\_\_ |
| «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г. | «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г. |

Москва, 2020

# 1 Цель работы

Построение основных графиков, входящих в этап разведочного анализа данных.

# 2 Ход работы

2.1 Выбор Dataset

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

2.2 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\dima\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

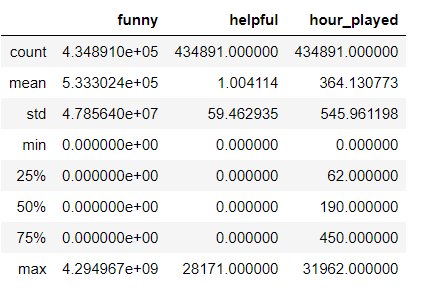
is\_early\_access\_review 0

recommendation 0

review 1516

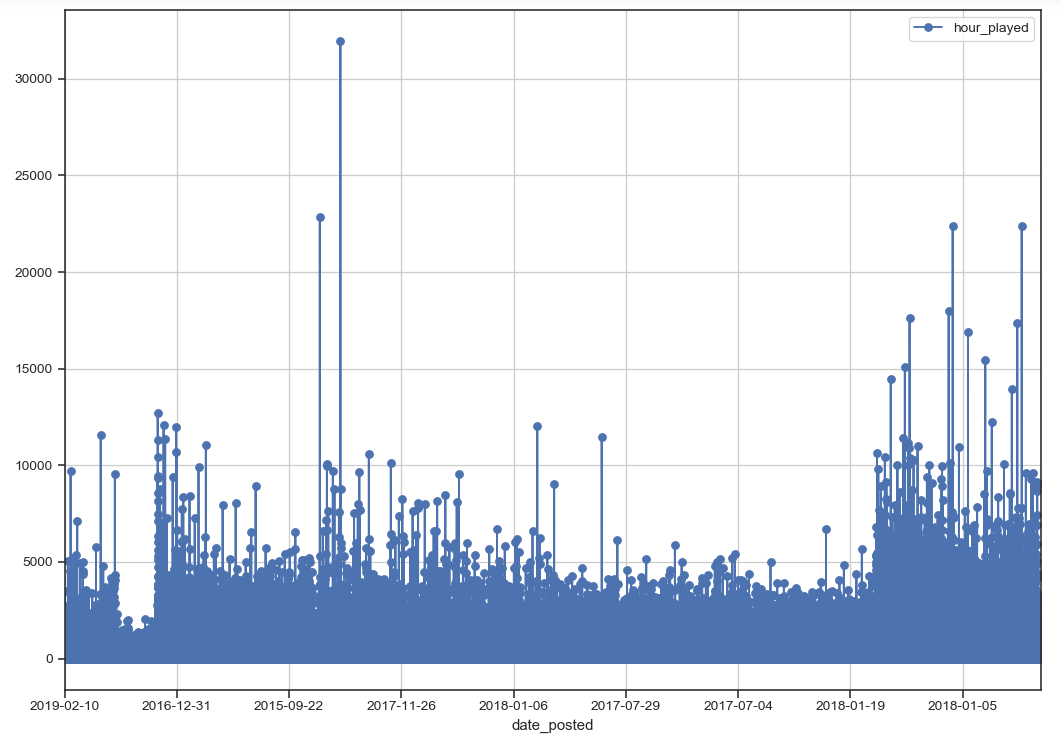
title 0

data.describe()

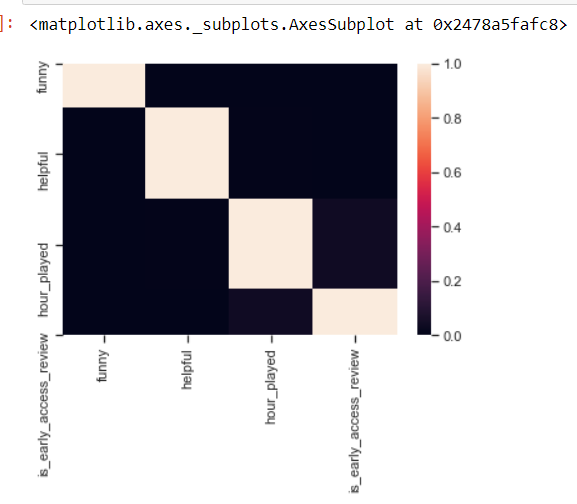


2.2 Графическая часть

Зависимость наигранных часов от даты поста



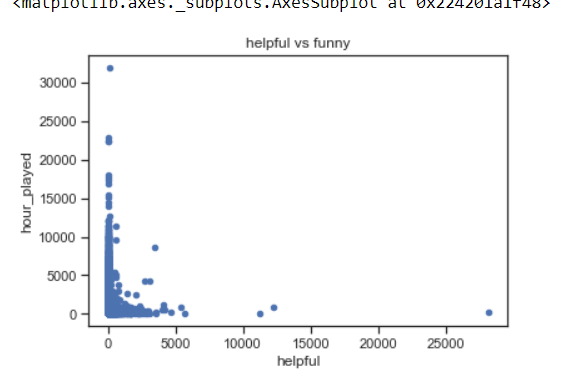
Тепловая карта



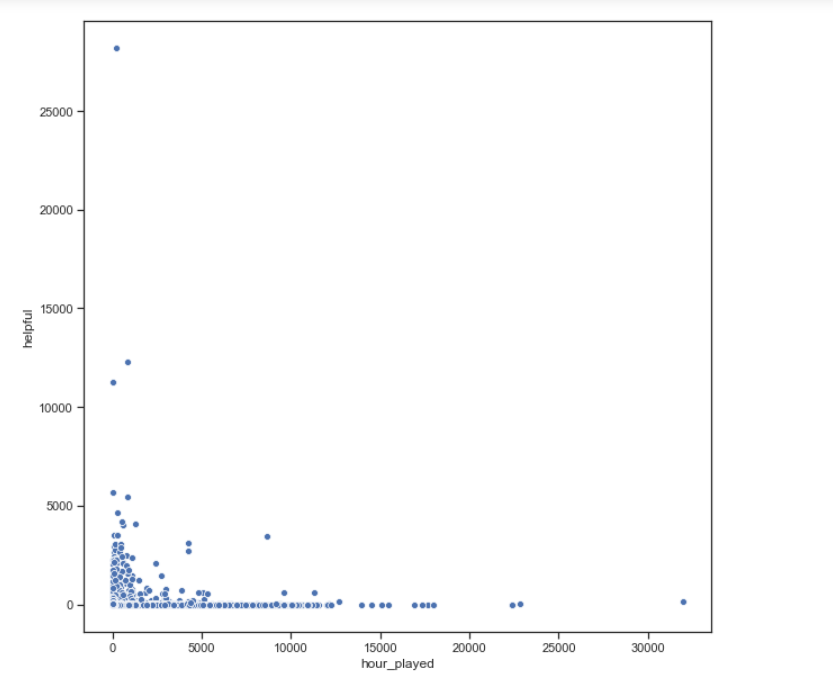
fig, ax = plt.subplots(figsize=(10,10))

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

