ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(национальный исследовательский университет)» (МАИ)

Направление подготовки: 27.03.05 «Инноватика»

**Лабораторная работа №2**

по дисциплине «ПМиАОЗД»

Выполнил:

Студенты гр. М3О-236Б-22

Попов М.А.

Кузьмин Д.Е.

Преподаватель:

Александрова С.С.

Москва 2023

Оглавление

[Цель работы 3](#_Toc153545657)

[Ход работы 3](#_Toc153545658)

[Вывод 6](#_Toc153545659)

# Цель работы

Анализ и обработка данных из открытого источника Kaggel, датасета Titanik при помощи библиотек pandas и numpy, а также построение графиков зависимости с помощью библиотеки matplotlib.pyplot. Построить три графика и написать к каждому графику пояснение кода.

# Ход работы

Задаём библиотеки и считываем файл данных:

**import** **pandas** **as** **pd**

**import** **numpy** **as** **np**

**import** **seaborn** **as** **sns**

**import** **matplotlib.pyplot** **as** **plt**

df = pd.read\_csv('train.csv')

df.head()

Закрепим passengerId как очновной индекс

df.set\_index('PassengerId').head()

Создадим круговую диаграмму для отображения количества мужчин и женщин на борту корабля

# суммируем количество мужчин и женщин

males = (df['Sex'] == 'male').sum()

females = (df['Sex'] == 'female').sum()

# помещяем их в лист, для дальнейшей работы с ними

proportions = [males, females]

# создание круговой диаграммы

plt.pie(proportions, labels = ['Males', 'Females'],shadow = True, colors = ['skyblue','orange'], explode = (**0.15** , **0**), startangle = **90**,autopct = '%1.1f%%')

plt.axis('equal')

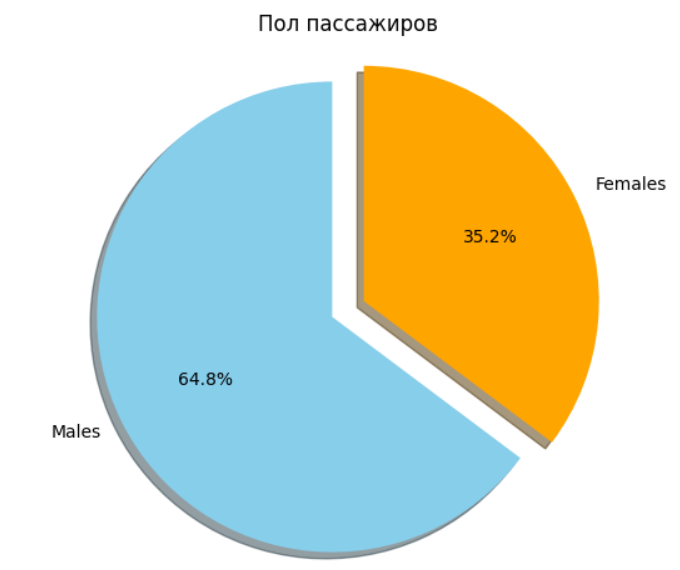
# задаём описание

plt.title("Пол пассажиров")

# jотображение

plt.tight\_layout()

plt.show()



Создадим диаграмму рассивания с указанием оплаченного тарифа и возраста пассажиров:

lm = sns.lmplot(x = 'Age', y = 'Fare', data = df, hue = 'Sex', fit\_reg=False)

# задаём описание

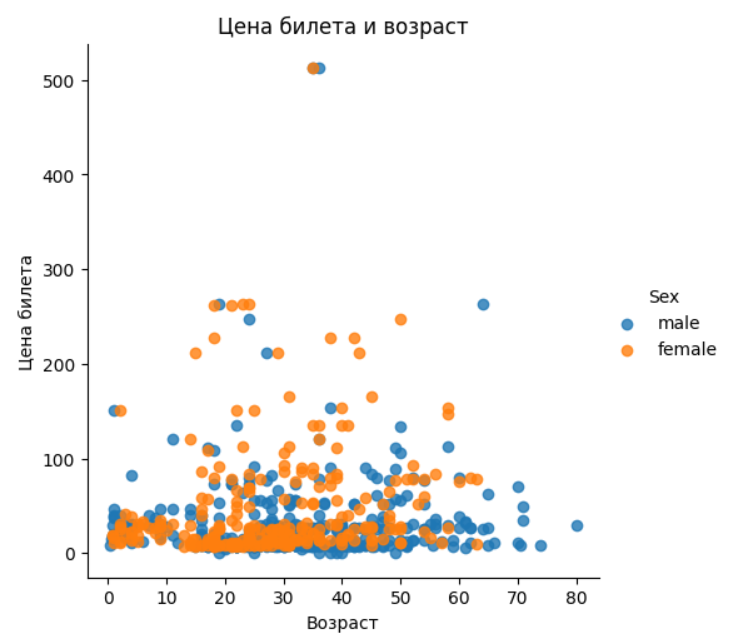
lm.set(title = 'Цена билета и возраст')

plt.xlabel('Возраст')

plt.ylabel('Цена билета')

axes = lm.axes

plt.show()



Посчитаем количество выживших человек:

df1 = df.Survived.sum()

df1

ОТВЕТ: 342

Создадим гистограмму для цены билета и зависимости от количества людей:

# сортируе значения от боьшего к меньшему по значению

dff = df.Fare.sort\_values(ascending = False)

dff

binsVal = np.arange(**0**,**600**,**10**)

plt.hist(dff, bins = binsVal, color='orange', alpha=**0.8**, align='mid', edgecolor='black', linewidth=**1**)

# задаём описание графика

plt.xlabel('Цена')

plt.ylabel('Количество')

plt.title('Плата за билет по кол-ву людей')

# отображаем график

plt.grid ( True )

plt.show()



# Вывод

Проведя анализ и обработку Датасета с информацией о пассажирах Титаника были построены графики, отображающие соотношение людей на борту, распределение людей по цене билета в зависимости от возраста и с указанием пола пассажира. Также построить гистограмму зависимости количества людей, купивших билет и цены этого билета.

Исходя из полученных данных, можно смело утверждать, что: мужчин на борту было больше почти в 2 раза, самые дорогие билеты были куплены мужчиной и женщиной и им было в районе 40 лет, самые дешёвые билеты были куплены в самом большом количестве.