Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования

«Ульяновский государственный технический университет»

Кафедра «Вычислительная техника»

Дисциплина «Системы искусственного интеллекта»

Лабораторная работа №1

«Анализ пассажиров корабля "Титаник"»

Выполнил:

студент группы ИВТАПбд-41

Сокольский Р.С

Проверил:

преподаватель кафедры «ВТ»

Святов К.В.

Ульяновск, 2025

Оглавление

[1. Постановка задачи 3](#__RefHeading___Toc165_3389309196)

[2. Реализация 4](#__RefHeading___Toc165_3389309196_%2525D0)

[3. Тестирование 8](#__RefHeading___Toc165_3389309196_%2525D1)

[4. Выводы 10](#__RefHeading___Toc165_3389309196_%2525D2)

# **1. Постановка задачи**

Необходимо дополнить программу, данную в условии, чтобы она могла определить:

1. Сколько мужчин / женщин было на борту?

2. Определите распределение функции Pclass. Теперь для мужчин и женщин отдельно. Сколько людей из второго класса было на борту?

3. Каковы медиана и стандартное отклонение Fare? Округлите до 2-х знаков после запятой.

4. Правда ли, что средний возраст выживших людей выше, чем у пассажиров, которые в конечном итоге умерли?

5. Это правда, что пассажиры моложе 30 лет. выжили чаще, чем те, кому больше 60 лет. Каковы доли выживших людей среди молодых и пожилых людей?

6. Правда ли, что женщины выживали чаще мужчин? Каковы доли выживших людей среди мужчин и женщин?

7. Какое имя наиболее популярно среди пассажиров мужского пола?

8. Как средний возраст мужчин / женщин зависит от Pclass? Выберите все правильные утверждения:

# **2****. Реализация**

Файл titanic\_train.csv содержит информацию о пассажирах.

С помощью библиотеки pandas был выполнен анализ информации.

Сначала мы загружаем данные:

**Листинг 1.** Загрузка данных

|  |
| --- |
| **import pandas as pd**  **data = pd.read\_csv("titanic\_train.csv")** |

Затем идут ответы на поставленные вопросы:

1. Сколько мужчин / женщин было на борту?

**Листинг 2.** Код ответа на вопрос

|  |  |
| --- | --- |
| |  | | --- | | **sex\_counts = data['Sex'].value\_counts()**  **print("Количество мужчин и женщин:")**  **print(sex\_counts)** |   Вывод:  Количество мужчин и женщин:  Sex  male 577  female 314  Name: count, dtype: int64 |

2. Определите распределение функции Pclass. Теперь для мужчин и женщин отдельно. Сколько людей из второго класса было на борту?

**Листинг 3.** Код ответа на вопрос

|  |  |
| --- | --- |
| |  | | --- | | **print("Распределение по классам для мужчин и женщин:")**  **sex\_and\_pclass = data.groupby(['Sex', 'Pclass']).size().unstack()**  **print(sex\_and\_pclass)**  **pclass2\_total = (data['Pclass'] == 2).sum()**  **print(f"\nКоличество людей из второго класса: {pclass2\_total}")** |   Вывод:  Распределение по классам для мужчин и женщин:  Pclass 1 2 3  Sex  female 94 76 144  male 122 108 347  Количество людей из второго класса: 184 |

3. Каковы медиана и стандартное отклонение Fare? Округлите до 2-х знаков после запятой.

**Листинг 4.** Код ответа на вопрос

|  |  |
| --- | --- |
| |  | | --- | | **median\_fare = round(data['Fare'].median(), 2)**  **std\_fare = round(data['Fare'].std(), 2)**  **print(f"Медиана Fare: {median\_fare}")**  **print(f"Стандартное отклонение Fare: {std\_fare}")** |   Вывод:  Медиана Fare: 14.45  Стандартное отклонение Fare: 49.69 |

4. Правда ли, что средний возраст выживших людей выше, чем у пассажиров, которые в конечном итоге умерли?

**Листинг 5.** Код ответа на вопрос

|  |  |
| --- | --- |
| |  | | --- | | **age\_survived = data[data['Survived'] == 1]['Age'].mean()**  **age\_not\_survived = data[data['Survived'] == 0]['Age'].mean()**  **print(f"Средний возраст выживших: {round(age\_survived, 2)}")**  **print(f"Средний возраст погибших: {round(age\_not\_survived, 2)}")**  **if age\_survived > age\_not\_survived:**  **print("Ответ: Да, средний возраст выживших выше, чем у погибших.")**  **else:**  **print("Ответ: Нет, средний возраст выживших не выше, чем у погибших.")** |   Вывод:  Средний возраст выживших: 28.34  Средний возраст погибших: 30.63  Ответ: Нет, средний возраст выживших не выше, чем у погибших. |

5. Это правда, что пассажиры моложе 30 лет. выжили чаще, чем те, кому больше 60 лет. Каковы доли выживших людей среди молодых и пожилых людей?

**Листинг 6.** Код ответа на вопрос

|  |  |
| --- | --- |
| |  | | --- | | **young = data[data['Age'] < 30]**  **old = data[data['Age'] > 60]**  **young\_survived = young['Survived'].mean() \* 100**  **old\_survived = old['Survived'].mean() \* 100**  **if young\_survived > old\_survived:**  **print("Да, пассажиры моложе 30 лет выживали чаще, чем те, кому больше 60 лет.")**  **else:**  **print("Нет, пассажиры моложе 30 лет выживали не чаще, чем те, кому было 60 лет.")**  **print(f"Доля выживших среди молодых (<30): {round(young\_survived, 1)} %")**  **print(f"Доля выживших среди пожилых (>60): {round(old\_survived, 1)} %")** |   Вывод:  Да, пассажиры моложе 30 лет выживали чаще, чем те, кому больше 60 лет.  Доля выживших среди молодых (<30): 40.6 %  Доля выживших среди пожилых (>60): 22.7 % |

6. Правда ли, что женщины выживали чаще мужчин? Каковы доли выживших людей среди мужчин и женщин?

**Листинг 7.** Код ответа на вопрос

|  |  |
| --- | --- |
| |  | | --- | | **survival\_by\_sex = data.groupby('Sex')['Survived'].mean() \* 100**  **if survival\_by\_sex['female'] > survival\_by\_sex['male']:**  **print("Да, женщины выживали чаще мужчин.")**  **else:**  **print("Нет, мужчины выживали чаще женщин.")**  **print(f"Доля выживших среди мужчин: {round(survival\_by\_sex['male'], 1)} %")**  **print(f"Доля выживших среди женщин: {round(survival\_by\_sex['female'], 1)} %")** |   Вывод:  Да, женщины выживали чаще мужчин.  Доля выживших среди мужчин: 18.9 %  Доля выживших среди женщин: 74.2 % |

7. Какое имя наиболее популярно среди пассажиров мужского пола?

* Charles
* Thomas
* William
* John

**Листинг 8.** Код ответа на вопрос

|  |
| --- |
| **male = data[data['Sex'] == 'male']['Name']**  **male\_names = male.str.extract(r'(?:Mr\.|Master\.|Rev\.|Dr\.|Major\.|Col\.|Don\.|Sir\.|Capt\.|Jonkheer\.)\s+([A-Za-z]+)', expand=False)**  **counts = male\_names.value\_counts()**  **top\_name = counts.idxmax()**  **top\_count = counts.max()**  **print(f"Самое популярное мужское имя: {top\_name} ({top\_count} раз).")** |

|  |
| --- |
| Вывод:  Самое популярное мужское имя: William (35 раз). |

8. Как средний возраст мужчин / женщин зависит от Pclass? Выберите все правильные утверждения:

* В среднем мужчины 1 класса старше 40 лет
* В среднем женщины 1 класса старше 40 лет
* Мужчины всех классов в среднем старше, чем женщины того же класса
* В среднем, пассажиры первого класса старше, чем пассажиры 2-го класса, которые старше, чем пассажиры 3-го класса.

**Листинг 9.** Код ответа на вопрос

|  |
| --- |
| **age\_by\_sex = data.groupby(['Sex', 'Pclass'])['Age'].mean()**  **age\_and\_class = data.groupby('Pclass')['Age'].mean()**  **print("\nСредний возраст пассажиров в зависимости от пола и класса:")**  **print(round(age\_by\_sex, 1))**  **print("\nВерные утверждения:")**  **# 1. В среднем мужчины 1 класса старше 40 лет**  **if age\_by\_sex['male', 1] > 40:**  **print("- В среднем мужчины 1 класса старше 40 лет")**  **# 2. В среднем женщины 1 класса старше 40 лет**  **if age\_by\_sex['female', 1] > 40:**  **print("- В среднем женщины 1 класса старше 40 лет")**  **# 3. Мужчины всех классов старше женщин того же класса**  **if all(age\_by\_sex['male', c] > age\_by\_sex['female', c] for c in [1, 2, 3]):**  **print("- Мужчины всех классов старше женщин того же класса")**  **# 4. Пассажиры 1 > 2 > 3 по возрасту**  **if age\_and\_class[1] > age\_and\_class[2] > age\_and\_class[3]:**  **print("- В среднем, пассажиры 1 класса старше 2-го, а те старше 3-го")** |

|  |
| --- |
| Вывод:  Средний возраст пассажиров в зависимости от пола и класса:  Sex Pclass  female 1 34.6  2 28.7  3 21.8  male 1 41.3  2 30.7  3 26.5  Name: Age, dtype: float64  Верные утверждения:  - В среднем мужчины 1 класса старше 40 лет  - Мужчины всех классов старше женщин того же класса  - В среднем, пассажиры 1 класса старше 2-го, а те старше 3-го |
|  |

# **3.** **Выводы**

В ходе выполнения лабораторной работе мы провели анализ данных о пассажирах Титаника используя библиотеку pandas в Python.