Міністерство освіти і науки України

Національний університет «Одеська політехніка»

Кафедра комп’ютеризованих систем та програмних технологій

Протокол лабораторної роботи №1

з дисципліни «Інтелектуальні системи»

на тему «БІБЛІОТЕКА ДЛЯ АНАЛІЗУ ДАНИХ PANDAS»

Виконав:

студент гр. АТ-191

Кинєв Є. О.

Перевірив:

д.т.н., проф. Фомiн О.О.

Одеса-2023

Мета роботи: отримання практичних навичок роботи з бібліотекою аналізу даних Pandas.

1. Для датасета Titanic (https://www.kaggle.com/datasets/heptapod/titanic) визначити стать та вік (з градацією у 10 років) пасажирів, які мали найбільшу ймовірність порятунку.

В якості структури зберігання даних використати бібліотеку Pandas.

2. Хід роботи:

2.1 Створити датафрейм Pandas та завантажити до нього датасет за допомогою функції read\_csv.

import pandas as pd

df = pd.read\_csv("./data.csv", index\_col=0)

2.2 Ознайомитись зі структурою даних датасета за допомогою функцій head, info, describe.

df.info(show\_counts=True)

#data

df.head()

#data

df.describe()

#data

2.3 Ознайомитись з розмірністю та типами даних датасета за допомогою атрибутів shape, dtypes.

df.shape

#data

df.dtypes

#data

2.4 Додати до датафрейму поле Age\_group як поле Age з градацією у 10 років.

mydf=df[['Sex','Age', 'Survived']]

mydf['Age\_group']=mydf['Age'].map(lambda x: int(x/10)+1)

mydf.head()

2.5 Розрахувати кількість пасажирів за кожною групою по полям 'Sex', 'Age\_group', 'Survived' за допомогою функції groupby.

mydf = mydf.groupby(['Sex', 'Age\_group', 'Survived'])

mydf = mydf.size().unstack()

mydf.head()

2.6 Візуалізувати отримані дані за допомогою функції plot.

mydf.plot(kind='barh', stacked=True)

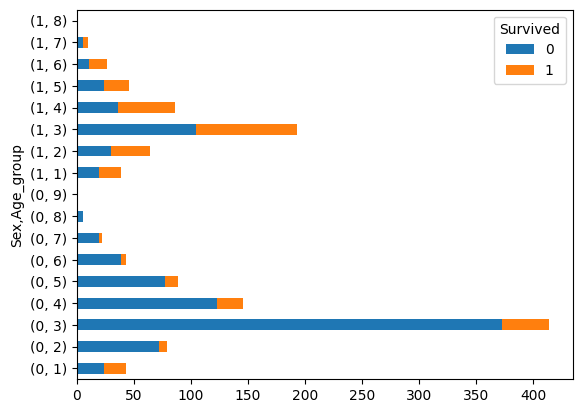


Рис. Графік врятованих

2.7 Висновки

Згідно з графіку можна побачити, що кількість чоловіків, які вижили було значно меньша ніж жінок. Також можна відстежити вікову групу учасників