

Практическая работа №5

Цель: Реализовать базовый анализ датасета Netflix, включая проверку на дубликаты, вывод информации о датасете, сортировку данных, удаление столбцов, замену пустых значений, а также сохранение обработанного датасета.

Ход работы:

1. Импорт библиотеки pandas:

```
import pandas as pd
```

2. Проверка на совпадения и их удаление:

```
dr = df.drop_duplicates()
```

3. Вывод информации о датасете:

```
print("Исходный размер датасета:", df.shape)
```

```
print("Размер датасета после удаления дубликатов:", dr.shape)
```

4. Вывод наименований колонок:

```
print("Наименования колонок:", df.columns.tolist())
```

5. Вывод уникальных значений для каждой колонки:

```
for column in df.columns:
```

```
    print(f"Уникальные значения для {column}:", df[column].nunique())
```

6. Отображение информации о датасете:

```
df.info()
```

7. Отображение статистического описания датасета:

```
df.describe(include='all')
```

8. Отсортировать датасет по определенным параметрам:

```
sorted_df = dr.sort_values(by='release_year', ascending=False)
```

9. Удаление ненужных столбцов:

```
reduced_df = sorted_df.drop(columns=['show_id'])
```

10. Замена пустых значений :

```
cleaned_df = reduced_df.fillna({'country': 'Unknown'})
```

11. Выборка данных в моём случае первых 10:

```
sampled_data = cleaned_df.loc[:9, ['title', 'release_year']]
```

12. Вывод примера выборки:

```
print(sampled_data)
```

13. Сохранение нового датасета:

```
newDataset = '/Users/oleg/Desktop/Питон ппактос/netflix_titles2.csv'  
cleaned_df.to_csv(newDataset, index=False)
```

Вывод: В результате анализа датасета Netflix были выполнены основные операции обработки данных, включая удаление дубликатов, сортировку, очистку и выборку данных. Эти шаги позволили подготовить датасет для дальнейшего анализа и использования в различных задачах анализа данных.