

day09_assignments

January 10, 2023

1 Assignments:

1. Columns in Titanic dataset

2. NaN values

3. Datasets in `sns.load_dataset()`

4. Export dataset in another directory

5. Import dataset from same directory * Author: Kaleem Ullah * Date: 2023-01-09 *
Email: kaleemrao417@outlook.com

1.1 1. Columns explanation in Titanic dataset

- **survived:**
How many people survived the titanic disaster(0 = No, 1 = Yes).
- **pclass:**
The class of the ticket the passenger purchased (1 = 1st, 2 = 2nd, 3 = 3rd).
- **sex:**
The passenger's gender (male or female).
- **age:**
The passenger's age in years.
- **sibsp:**
The number of siblings or spouses the passenger had aboard the Titanic.
- **parch:**
The number of parents or children the passenger had aboard the Titanic.
- **fare:**
The fare the passenger paid.
- **embarked:**
The port where the passenger embarked (C = Cherbourg, Q = Queenstown, S = Southampton).
- **class:**
The class of the ticket the passenger purchased (1 = 1st, 2 = 2nd, 3 = 3rd).
- **who:**
The passenger is man, woman or child.
- **adult_male:**
The passenger is adult male or not(True = Yes, False = No).
- **deck:**
The deck of the Titanic where the passenger was staying.

- **embark_town:**
The town from where the passenger embarked.
- **alive:**
The passenger survived or not.
- **alone:**
The passenger was alone or not.

1.2 2. What are NaN's and what are their importance in dataset?

- NaN stands for **Not a Number**.
It is a special value that is used to represent missing values in a dataset.
- It is important to handle missing values in a dataset because most of the machine learning algorithms do not support missing values.

1.3 3. How many datasets are in sns.load__dataset() and how to fetch?

```
print(len(sns.get_dataset_names()))
print((sns.get_dataset_names()))
sns.load_dataset('healthexp')
```

1.4 4. Export dataset in another directory

```
kashti.to_csv('../day_8/titanic.csv')
```

1.5 5. Load dataset from same directory

```
df = pd.read_csv('titanic.csv')
df.head()
```