EXPLORATORY DATA ANALYSIS CHEAT SHEET: SEABORN, PANDAS, MISSINGO

Classification

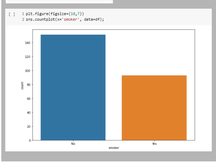
-Check unique values target column

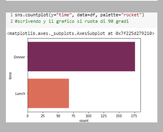
```
pd.unique(df['nomecolonna'])
```

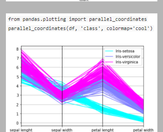
-Check if problem is balanced

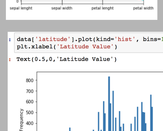
```
df.groupby('class').count()
```

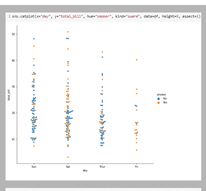
```
df['Target'].value_counts()
```

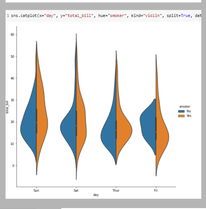


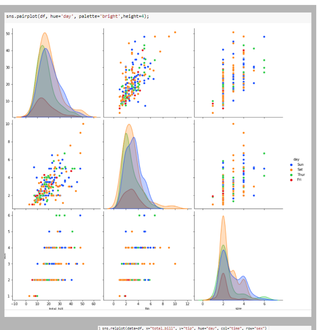


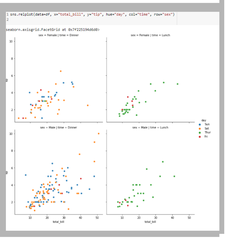


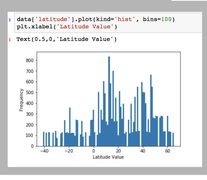






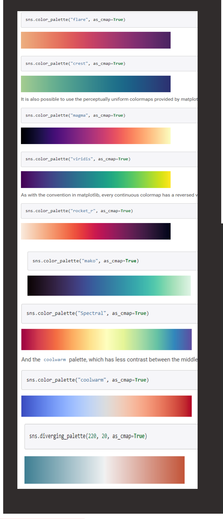


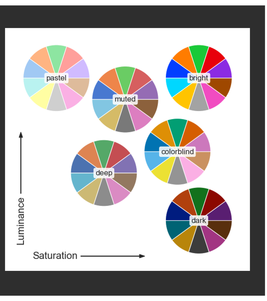




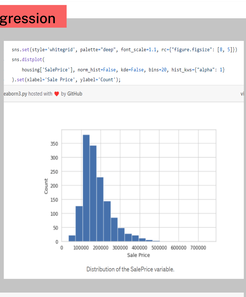
Color Palette

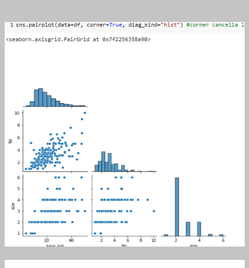
It's important to emphasize here that using red and green, while intuitive, **should be avoided.**

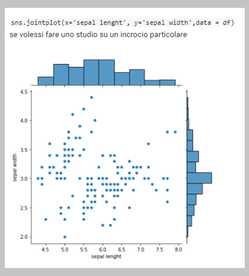


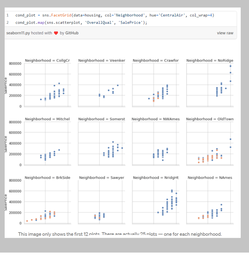


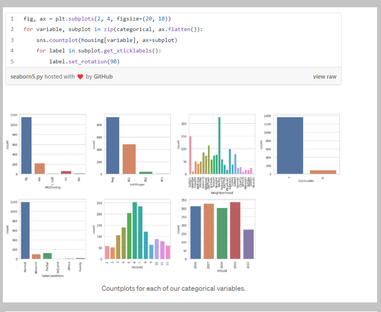
Regression

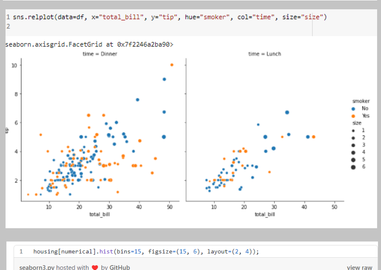






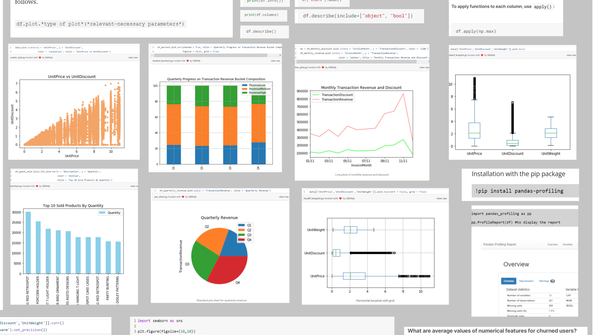


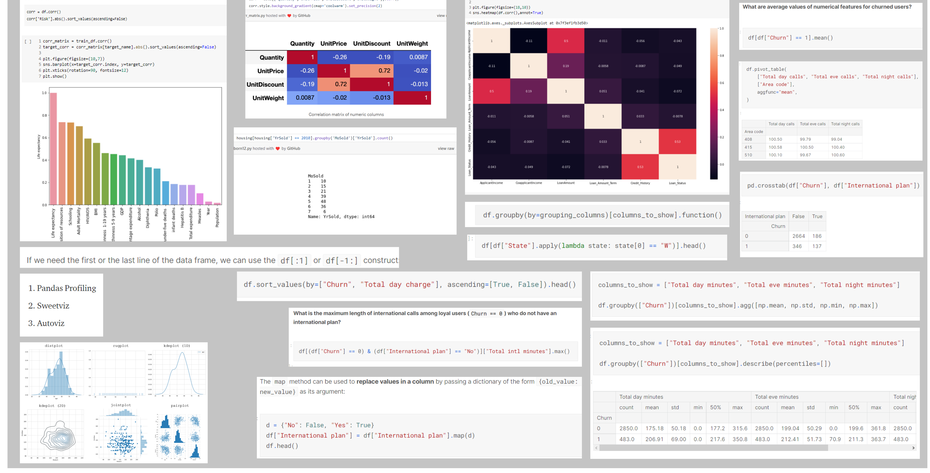




Missingo

Other pandas eda





```
df.groupby(by=grouping_columns)[columns_to_show].function()
```

```
df[df['State'].apply(lambda state: state[0] == 'W')].head()
```

If we need the first or the last line of the data frame, we can use the `df[:1]` or `df[-1:]` construct

1. Pandas Profiling
2. Sweetviz
3. Autoviz

```
df.sort_values(by=['Churn', 'Total day charge'], ascending=[True, False]).head()
```

What is the maximum length of international calls among loyal users (`Churn == 0`) who do not have an international plan?

```
df[(df['Churn'] == 0) & (df['International plan'] == 'No')]['Total intl minutes'].max()
```

```
columns_to_show = ['Total day minutes', 'Total eve minutes', 'Total night minutes']

df.groupby(['Churn'])[columns_to_show].agg([np.mean, np.std, np.min, np.max])
```

```
columns_to_show = ['Total day minutes', 'Total eve minutes', 'Total night minutes']

df.groupby(['Churn'])[columns_to_show].describe(percentiles=[])
```

The `map` method can be used to **replace values in a column** by passing a dictionary of the form `{old_value: new_value}` as its argument:

```
d = {'No': False, 'Yes': True}
df['International plan'] = df['International plan'].map(d)
df.head()
```