___dataprep

April 2, 2024

```
[ ]: ls
```

```python
[ ]: import pandas as pd
```

```python
[ ]: df = pd.read_csv("alim.csv")
```

```python
[ ]: df
```

```python
[ ]: df.Produit
```

```python
[ ]: df.Produit.unique()
```

```python
[ ]: df.Produit = df.Produit.str.lower()
     df
```

```python
[ ]: produits = list(df.Produit.unique())
```

```python
[ ]: produits = [i.lower() for i in produits]
     produits
```

```python
[ ]: [i for i in produits if "chiken" in i]
```

```python
[ ]: [i for i in produits if "poulet" in i]
```

```python
[ ]: [i for i in produits if "volai" in i]
```

```python
[ ]: df = df.loc[df.Produit == "viande de volailles"]
     df
```

```python
[ ]: df.nunique()
```

```python
[ ]: tmp = df.nunique()
     tmp = tmp[tmp > 1]
     tmp
```

```python
[ ]: tmp.index
```

```python
df = df.loc[:, tmp.index]
df
```

```python
dd = {}
for i, ser in df.iterrows():

    k, v = ser["Élément"], ser["Unité"]
    dd[k] = v
dd
```

```python
cols = ["Code zone", "Zone", "Élément", "Valeur"]
cols
```

```python
df = df.loc[:, cols]
df
```

```python
df = df.pivot(index=["Code zone", "Zone"], columns="Élément", values="Valeur")
df
```

```python
df = df.reset_index()
df.columns.name = None
df
```

```python
nan_values = df.isna().mean().round(2)
nan_values
```

```python
nan_values.sort_values(ascending=False)
```

```python
cols = [
    "Code zone",
    "Zone",
    "Disponibilité intérieure",
    "Importations - Quantité",
    "Disponibilité de protéines en quantité (g/personne/jour)",
    "Disponibilité alimentaire (Kcal/personne/jour)",
    "Exportations - Quantité",
    "Résidus",
    "Variation de stock",
    "Production",
    "Nourriture",
]


df = df.loc[:, cols]
df
```

```python
df.isna().mean().round(2)
```

```python
df = df.fillna(0)
df
```

```python
df.isna().mean().round(2)
```

```python

```

```python
cols = [
    "code_zone",
    "zone",
    "dispo_int",
    "import",
    "dispo_prot",
    "dispo_alim",
    "export",
    "residus",
    "var_stock",
    "prod",
    "nourriture",
]
df.columns = cols
df
```

```python
corr = df.select_dtypes(float).corr()
corr
```

```python
import numpy as np
import seaborn as sns

mask = np.triu(corr)
sns.heatmap(corr, annot=True, mask=mask, vmax=1, vmin=-1, cmap="coolwarm",
    fmt=".2f")
```

```python
pop = pd.read_csv("pop.csv")
pop
```

```python
pop = pop.pivot(index=["Code zone", "Zone"], columns="Année", values="Valeur")
pop.reset_index(inplace=True)
pop.columns.name = None
pop
```

```python
pop.isna().mean().round(2)
```

```python
pop.columns
```

```python
cols = [
    "Code zone",
```

```
    "Zone",
    2017,
]
pop = pop.loc[:, cols]
pop
```

```
cols = [
    "code_zone",
    "zone",
    "population",
]
pop.columns = cols
pop
```

```
pop = pop.loc[:, ["code_zone", "population"]]
pop
```

```
df.merge(pop, on="code_zone", how="left")
df = df.merge(pop, on="code_zone", how="left")
df
```

```
df.to_csv("chicken.csv", index=False)
```

[ ]: