AnalisisyCuracion_Parte-II_Curation

August 13, 2020

0.1 ANALISIS DE TENDENCIAS DE CONSUMOS E INFRAESTRUCTURAS

By JoseForguez & DiegoTondo, 2020-08-09

[ ]:
```
import pandas as pd
import numpy as np
import warnings
import glob
import zipfile
import seaborn as sns
import matplotlib.pyplot as plt
from datetime import datetime

import plotly.graph_objs as go
import plotly.offline as plotly
from plotly.subplots import make_subplots

import matplotlib.dates as md
```

[ ]:
```
#pd.set_option('display.max_columns', 50)
```

LOAD TRANSACTIONS

[ ]:
```
df_transa = pd.read_csv('../dataset/data_csv/sis_transa_201801_202007.csv',␣
↪parse_dates=['fecha'])
```

[ ]:
```
df_transa.columns
```

[ ]:
```
c2drop = ['acum_usuario',
          'acum_vehiculo',
          'campovar1',
          'campovar2',
          'km_transaccion',
          'local_price',
          'odometro',
          'tag_description_1',
          'tag_description_2',
          'user_tag_id',
          'valor',
          'veh_efficiency'
```

```
        ]
```

[ ]:
```
df_transa.drop(columns=c2drop, axis=1, inplace=True)
```

[ ]:
```
df_transa[:5]
```

[ ]:
```
df_transa[pd.isna(df_transa['cantidad'])]
```

[ ]:
```
df_transa.dropna(subset=['cantidad'], inplace=True)
```

[ ]:
```
len(df_transa[df_transa['cantidad'].astype('string').str.contains(',')])
```

[ ]:
```
df_transa['cantidad'] = df_transa['cantidad'].astype('string').str.replace(',',␣
↪'.').astype('float64')
```

[ ]:
```
df_transa.dtypes
```

[ ]:
```
df_transa.shape
```

[ ]:
```
df_transa = df_transa[df_transa['cantidad']>=0]
df_transa.shape
```

[ ]:
```
df_transa = df_transa[~(df_transa['fecha']<'2018-01-01')]
df_transa.shape
```

[ ]:
```
df_transa[df_transa['id_equipo'].isna()]
```

[ ]:
```
df_transa[['id_vehiculo', 'cantidad']]
```

LOAD SITES

[ ]:
```
df_equipo = pd.read_csv('../dataset/data_csv/fs_equipo.csv')
df_equipo.columns
```

[ ]:
```
df_equipo.drop(axis=1, columns=['direccion_ip', 'pass', 'sync', 'online'],␣
↪inplace=True)
```

[ ]:
```
df_equipo.shape
```

[ ]:
```

```

LOAD COMPANIES

[ ]:
```
df_emp = pd.read_csv('../dataset/data_csv/fs_empresa_tagged.csv')
```

[ ]:
```
df_emp.columns
```

[ ]:
```
df_emp.drop(axis=1, columns=['cuit', 'telefono', 'email', 'pais', 'id_pais',␣
↪'provincia', 'ciudad', 'direccion', 'cp', 'ultima_fecha_sync_db'],␣
↪inplace=True)
```

[ ]:
```
df_emp = df_emp.convert_dtypes()
df_emp.dtypes
```

[ ]:
```

```

LOAD VEHICULES

[ ]:
```
df_veh = pd.read_csv('../dataset/data_csv/fs_vehiculos.csv')
```

[ ]:
```
df_veh.columns
```

[ ]:
```
df_veh
```

[ ]:
```
df_veh.drop(df_veh.loc[:,'id_equipo':'sync'], axis=1, inplace=True)
```

[ ]:
```
df_veh
```

[ ]:
```

```

Merge Dataframes

[ ]:
```
df_equipo.shape, df_emp.shape
```

[ ]:
```
dfe = pd.merge(df_equipo, df_emp, on='id_empresa')
dfe.shape
```

[ ]:
```
df_transa.shape
```

[ ]:
```
df = pd.merge(df_transa, dfe, on='id_equipo', how='left')
df.shape
```

[ ]:
```
df = df[df['baja']==0]
```

[ ]:
```
df.shape
```

[ ]:
```
dfa = pd.merge(df, df_veh, on='id_vehiculo', how='left')
dfa.shape
```

[ ]:
```
dfa[['fecha', 'id_vehiculo', 'main_id', 'id_empresa', 'cantidad', 'segmento',␣
↪'id_tanque', 'id_bomba', 'empresa', 'descripcion']]
```

[ ]:
```
list(dfa[dfa['id_empresa'].isnull()]['id_equipo'].unique())
```

[ ]:
```
dfa = dfa.dropna(axis=0, subset=['id_empresa'])
dfa.shape
```

[ ]:
```

```

[ ]:
```
dfa[:5]
```

[ ]:
```

```

[ ]:
```
#dfa.to_csv('../dataset/data_csv/sis_transa_201801_202007_merged.csv',␣
↪index=False)
```