



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



1º PROJETO: ANÁLISE EXPLORATÓRIA DE DADOS EM PYTHON II



GUIA DA AULA 2



Visualize dados



Acompanhe aqui
os temas que
serão tratados
na videoaula

● **Mapa de
entregas
por região**

● **Gráfico de
entregas
por região**



1. Mapa de entregas por região

Vamos utilizar o pacote Python Geopandas para visualizar as coordenadas dos **hubs** e das **entregas** no mapa do Distrito Federal, segmentados pela região dos hubs. O pacote adiciona funcionalidades geoespaciais ao pacote Python Pandas.

A documentação do pacote Geopandas pode ser encontrada no *link*:

<https://geopandas.org/>.

```
In [ ]: !pip3 install geopandas ;
```

```
In [ ]: import geopandas
```



• Mapa do Distrito Federal

Vamos fazer o *download* dos dados do mapa do Distrito Federal do *site* do IBGE no *link*

<https://www.ibge.gov.br/geociencias/cartas-e-mapas/bases-cartograficas-continuas>

para criar o DataFrame `mapa`. Note a coluna `geometry`:

```
In [ ]: !wget -q << EOF
https://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas /
bc100/go_df/versao2016/shapefile/bc100_go_df_shp.zip
EOF \ -O distrito-federal.zip
!unzip -q distrito-federal.zip -d ./maps
!cp ./maps/LIM_Unidade_Federacao_A.shp ./distrito-federal.shp
!cp ./maps/LIM_Unidade_Federacao_A.shx ./distrito-federal.shx
```

```
In [ ]: mapa = geopandas.read_file("distrito-federal.shp")
mapa = mapa.loc[[0]]
mapa.head()
```



• Mapa dos Hubs

Vamos criar o DataFrame `geo_hub_df` através do DataFrame `deliveries_df`.
 Note a nova coluna `geometry`:

```

In [ ]: hub_df = deliveries_df[["region", "hub_lng", "hub_lat"]].drop_duplicates (
        ).reset_index(drop=True)
        geo_hub_df = geopandas.GeoDataFrame (
            hub_df,
            geometry=geopandas.points_from_xy (
                hub_df["hub_lng"],
                hub_df["hub_lat"]
            )
        )
        geo_hub_df.head()
  
```



• Mapa das entregas

Vamos criar o DataFrame `geo_deliveries_df` através do DataFrame `deliveries_df`. Note a nova coluna `geometry`:

```

In [ ]: geo_deliveries_df = geopandas.GeoDataFrame (
        deliveries_df,
        geometry=geopandas.points_from_xy (
            deliveries_df ["delivery_lng"],
            deliveries_df ["delivery_lat"]
        )
    )
    geo_deliveries_df.head()
  
```



• Visualização

```

In [ ]: import matplotlib.pyplot as plt

# cria o plot vazio
fig, ax = plt.subplots(figsize = (50/2.54, 50/2.54))

# plot mapa do distrito federal
mapa.plot(ax=ax, alpha=0.4, color="lightgrey")

# plot das entregas
geo_deliveries_df.query("region == 'df-0'").plot(
    ax=ax, markersize=1, color="red", label="df-0"
)
geo_deliveries_df.query("region == 'df-1'").plot(
    ax=ax, markersize=1, color="blue", label="df-1"
)
geo_deliveries_df.query("region == 'df-2'").plot(
    ax=ax, markersize=1, color="seagreen", label="df-2"
)
  
```




```

# plot dos hubs
geo_hub_df.plot(
    ax=ax, markersize=30, marker="x", color="black", label="hub"
)

# plot da legenda
plt.title(
    "Entregas no Distrito Federal por Região" ,
    fontdict={"fontsize": 16}
)
lgnd = plt.legend(prop={"size": 15})
for handle in lgnd.legendHandles:
    handle.set_sizes([50])
  
```

- **Insights:**

1. As **entregas** estão corretamente alocadas aos seus respectivos **hubs**;
2. Os **hubs** das regiões 0 e 2 fazem **entregas** em locais distantes do centro e entre si, o que pode gerar um tempo e preço de entrega maior.



2. Gráfico de entregas por região

In []:

```

data = pd.DataFrame (
    deliveries_df [['region',
        'vehicle_capacity' ]].value_counts (normalize=True
    ).reset_index ()
data.rename (columns={0: "region_percent" }, inplace=True)
data.head ()

```



• Visualização

```

In [ ]: import seaborn as sns

with sns.axes_style('whitegrid'):
    grafico = sns.barplot(
        data=data,
        x="region",
        y="region_percent",
        ci=None,
        palette="pastel"
    )
    grafico.set(
        title='Proporção de entregas por região' ,
        xlabel='Região',
        ylabel='Proporção'
    );
  
```



- **Insights:**

1. A distribuição das entregas está muito concentrada nos hubs das regiões 1 e 2, mas pouco no da região 0. Contudo a capacidade dos veículos é a mesma para todos os hubs, logo os veículos poderiam ser deslocados para as regiões de maior tráfego.

