

Kriminalitetsanalyse

Installer biblioteker

```
In [ ]: !pip install pandas geopandas shapely numpy matplotlib mapclassify pyarrow
```

Indlæs biblioteker

```
In [2]: import pandas as pd
import geopandas as gpd
import numpy as np
import matplotlib.pyplot as plt
from shapely.geometry import box, Polygon
```

Indlæs data

```
In [21]: # Indlæs data
df = pd.read_csv("data/crimes_chi.csv")
```

```
In [ ]: df.head()
```

```
In [26]: # Indlæs data
date_format = "%m/%d/%Y %I:%M:%S %p"

df = pd.read_csv(
    "data/crimes_chi.csv",
    parse_dates=["Date"],
    date_format="%m/%d/%Y %I:%M:%S %p"
)
```

```
In [ ]: df.head()
```

```
In [30]: df_2024 = df[df["Date"].dt.year == 2024]
```

```
In [ ]: df_2024.head()
```

```
In [ ]: df_2024.describe()
```

Check data

```
In [40]: assert {"Longitude", "Latitude"}.issubset(df_2024.columns), "CSV skal have
```

```
In [41]: # Projicér til metrisk CRS for Danmark (ETRS89 / UTM 32N)
gdf = gpd.GeoDataFrame(
    df_2024,
    geometry=gpd.points_from_xy(df_2024.Longitude, df_2024.Latitude),
    crs="EPSG:4326"
).to_crs("EPSG:25832")
```

Gem arbejdsdata

```
In [ ]: # Gem et effektivt arbejdsformat til senere brug
gdf.to_parquet("data/crime_incidents_utm.parquet", index=False)
len(gdf), gdf.crs
```

Analyse 1: Beskrivende

```
In [ ]: df_desc = df_2024.groupby([df["Date"].dt.month, 'Description']).agg({
    'Description': 'count',
    'Arrest': 'sum'
}).rename(columns={'Description': 'Total_Incidents'})

print(df_desc)
```

```
In [73]: df_desc['Pct_Arrests'] = (df_desc['Arrest'] / df_desc['Total_Incidents'])
```

```
In [ ]: print(df_desc)
```

Analyse 2: Kortlægning

```
In [48]: # Lav et kvadratgitter
cell = 2000 # meter x meter
minx, miny, maxx, maxy = gdf.total_bounds
buf = cell
minx -= buf; miny -= buf; maxx += buf; maxy += buf

xs = np.arange(minx, maxx + cell, cell)
ys = np.arange(miny, maxy + cell, cell)

polys = [box(x, y, x+cell, y+cell) for x in xs[:-1] for y in ys[:-1]]
grid = gpd.GeoDataFrame(geometry=polys, crs=gdf.crs)
```

```
In [ ]: grid.plot(edgecolor="black", facecolor="none", figsize=(6,6))
gdf.plot(ax=plt.gca(), color="red", markersize=2)

plt.show()
```

```
In [50]: # Spatial join: tæl hændelser pr. celle
joined = gpd.sjoin(gdf[["geometry"]], grid, predicate="within", how="left")
counts = joined.groupby("index_right").size()
grid["count"] = counts.reindex(grid.index, fill_value=0)
```

```
In [ ]: # Plot
ax = grid.plot(column="count", scheme="Quantiles", k=10, legend=True, fig
plt.title("Hændelser pr. {} m celle (kvantiler)".format(cell))
```

```
plt.tight_layout()  
plt.show()
```

In []:

In []:

In []:

In []: