



Mapping Urban Vulnerability areas
(Crimes, Disasters, etc.) using Open Source Data

OMDENA SOUTH AFRICA

DATA COLLECTION

TASK-1



Initial Investigation

CRIME STASTICS

Kaggle - Data sets were older than the requirements

SAPS - Data sets were more up to date but were in password protected workbooks so proved difficult to extract

Supplimentary Data - This has proved very difficult as data sets are either pay per use or via application

IMAGES

Google - This was the best source of images and several scripts were written by team members to scrape the images

Supplied by Sheneese - Crime images already supplied at start of project

Data Collection

STATICAL INFORMATION

Interviews

We managed to extract data from SAP's excel files and have data from about 2016 to 21 in monthly format.



We added two extra features, latitude and longitude.

Focus Groups

IMAGERY



DATA

COLLECTION TOOLS



Observation

We have a great team working on collecting images.

Image sorting and resizing for the Model team.

Usage Data

Pre processing was carried out and images were resized to 300x300.



Surveys



DATA COLLECTED

3,347,784 rows

Over 2000 images

Challenges

IMAGE SOURCING

Time - Large variety of imagery required.

IMAGE RESIZING

Deciding on the best way to do this.

DATA COLLECTION

The sheer lack of publically available data on South Africa.

```
2 ...
3 A function to compare average holiday R
4 average daily counts with percentage ri
5 ...
6 Argu
7 name (str): Name of the holiday, fo
8 end (int): End date of the holiday pe
9 yr (int): year (range: 2005 - 2010)
10 days (int): Length of holiday period
11 svty (str): Severity (Slight, Serio
12 ...
13 ...
14 holiday = len(
15 df_accident[
16 ((df_accident['accident_severit
17 & (df_accident['date'] >= start
18 & ((df_accident['date'] <= end)
19 ) / days
20 ...
21 severity = len(
22 df_accident[
23 ((df_accident['accident_severit
24 & (df_accident['year'] == yr))
25 ...
26 ) / 365
27 perc = (int(holiday) - int(severity)) /
28 ...
29 print(
30     'The average', svty, 'RTCs over', n
31     int(holiday), 'per day.\nThe daily
32     int(severity), 'fatal accidents per
33 )
```

DATA PREPROCESSING

```

24     <b>Vehicles: </b> %s <br />
25     <b>Date: </b> %s
26     ...
27     map = flm.Map(location=[lat[1], lon[1]], zoom_start=12, scrollWheelZoom=False)
28
29     fg = flm.FeatureGroup(name='My V Map')
30
31     for lt, ln, nm, st, ca, ve, da in zip((lat), (lon), (nm), (st), (ca), (ve), (da)):
32         iframe = flm.IFrame(html = html % ((nm), (st), (lt), (ln), (ca), (ve), (da)), height = 165)
33         popup = flm.Popup(iframe, min_width=200, max_width=500)
34         fg.add_child(flm.CircleMarker(location = [lt, ln], popup = (popup), fill_color=color_producer(st), color='None', radius=10, fill_opacity = 0.7))
35     map.add_child(fg)
36
37     map.save('C:/Users/kmedr/My Drive/Omdena_Predicting_RTC_main/reports/map.html')
38     return map

```

Sample Code



```

1 df = pd.read_csv('/kaggle/input/crime-rate/crime_rate_2006-2016.csv')
2 df.head()
3
4 # We will correct the name of all provinces in order to make the merge smoothly
5 df["Province"].replace('Kwazulu/Natal','KwaZulu-Natal',inplace=True)
6
7 # we will calculate the total crimes from the year 2005 to 2016
8 m=list(df.columns)[8:19]
9 df['Total'] = df[m].sum(axis=1)
10 df['Total']
11
12 # We will calculate the Total crimes grouped by provinces
13 df_province=df.groupby("Province")["Total"].sum().sort_values(ascending=False).reset_index()
14 df_province
15
16 province_map = gpd.read_file("/kaggle/input/south-africa-government-boundaries/zaf_adm_sadb_ocha_20201109_SHP/zaf_admbnda_adm1_sadb_ocha_20201109.dbf")
17
18 province_map.rename(columns = {'ADM1_EN':'Province'}, inplace = True)
19
20 province_map1=province_map[["Province","geometry"]]
21
22 df_merged = pd.merge (province_map1, df_province, on= ['Province'], how='inner')
23
24 # We can visualize the total crimes per province along the map of South Africa
25 df_merged.explore(column='Total')
26

```

Geospatial



```

1 def get_weather(district, lat, lon):
2     api_address = ("https://archive-api.open-meteo.com/v1/era5?"
3                     f"latitude={lat}&"
4                     f"longitude={lon}&"
5                     "start_date=2016-01-01&"
6                     "end_date=2021-12-31&"
7                     "daily=temperature_2m_max,temperature_2m_min,precipitation_sum,rain_sum,windspeed_10m_max&"
8                     "timezone=Africa/Johannesburg")
9     json_data = requests.get(api_address).json()
10    data = pd.DataFrame(json_data['daily'])
11    data['latitude'] = lat
12    data['longitude'] = lon
13    data['district'] = district
14
15    return data

```

Weather Extraction

Types of Data

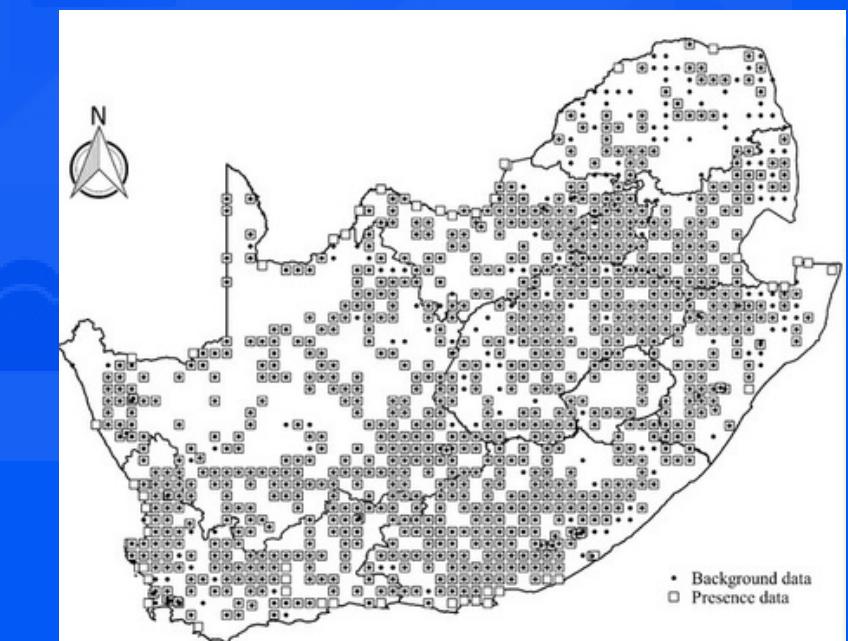
DATA SOURCES

Statistical:

- 'station' - Police station where crime was reported
- 'province' - The province where the crime was committed
- 'district' - The district where the crime was committed
- 'crime_category' - The type of crime committed
- 'date' - The month the crime was committed in
- 'number_of_crimes' - The number of crimes committed in the month

We then collected latitude and longitude of the stations and added two more feature:

- 'latitude'
- 'longitude'

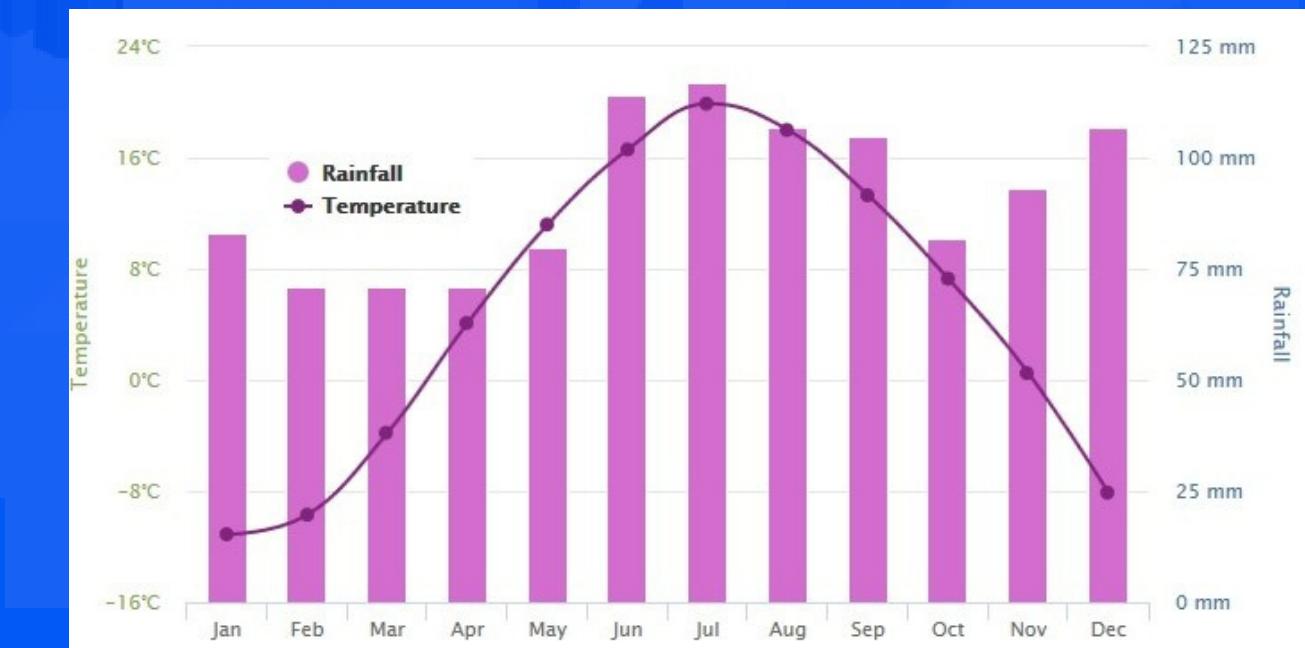


Types of Data

DATA SOURCES

Weather Data:

- 'district'
- 'date'
- 'temperature_2m_max'
- 'temperature_2m_min'
- 'precipitation_sum'
- 'rain_sum'
- 'windspeed_10m_max'

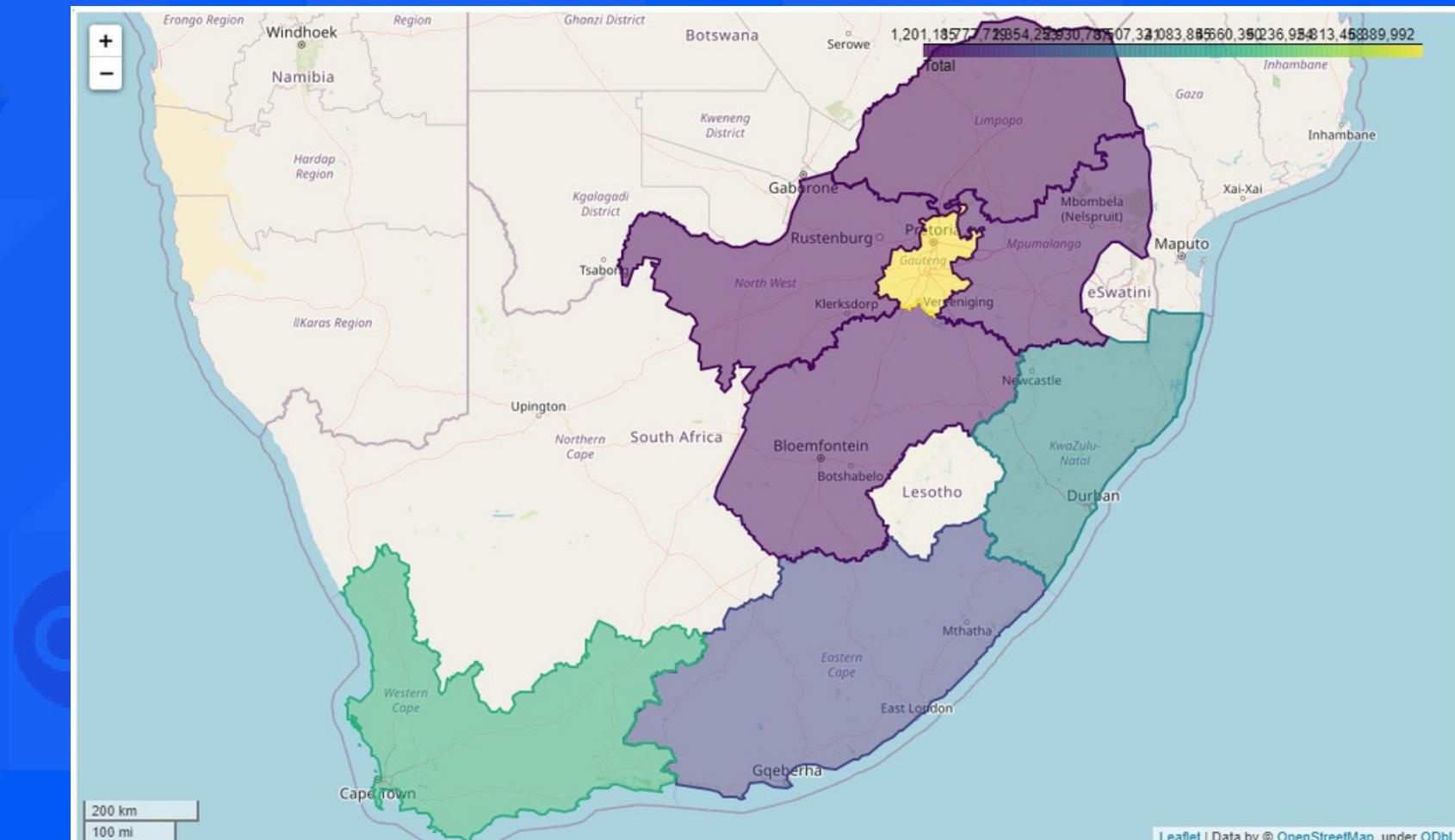


Types of Data

DATA SOURCES

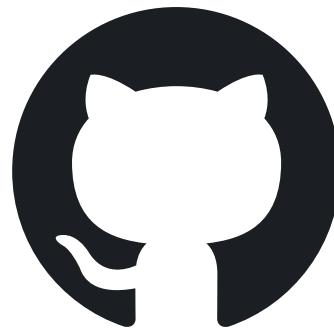
Geospatial:

- 'province'
- 'station'



Useful Links

LINKS



GitHub

<https://github.com/OmdenaAI/south-africa-chapter-mapping-urban-v>



Notion

<https://www.notion.so/South-Africa-Chapter-Mapping-Urban-Vulnerability-areas-Crimes-Disasters-etc-using-Open-Source-69e24d6f5c04497281afb3e8d742cbde>



Omdena

<https://omdena.com/challenges/mapping-urban-vulnerability-areas-using-open-source-data/>

Special Thanks To...

Statistical

- Kevin Medri
- Biswajit Mondal
- Ntandoyenkosi Matshisela
- Chukwudi Okereafor
- Hesham Saber
- Nikhila Radhaddevi
- Hesham Saber

Images

- Sheneese Bhoola
- Nikhila Radhaddevi
- Chukwudi Okereafor
- Biswajit Mondal
- Muhammad Rizwan Ali
- Swati Kashyap