# Mini-Project 1: Data Analysis & Visualization

## Dataset description

### Finnish bird nesting Atlas:

<https://www.avoindata.fi/data/en/dataset/lintuatlas>

The dataset when downloaded provides 3 csv-files and instructions on the contents of it inside a txt-file.

The information inside lajit.csv is as follows: an abbreviated name for the bird species, scientific name, Finnish name, Swedish name, English name and a binary mark if the bird’s information is public.

ruudut.csv contains location information which I discarded for this assessment.

havainnot.csv has information on each sighting from the atlases 1974-97 publication, 1986-89 publication and the combined sighting information. Sightings information contains: coordinates, nesting certainty number by year and combined and the species abbreviated.

the file ohje.txt contains the necessary information to map the appropriate full names to the abbreviations and numbers.

### Visits at the City of Vantaa swimming pools:

<https://www.avoindata.fi/data/en/dataset/vantaan-kaupungin-uimahallien-kaikki-kayntitapahtumat>

All data contained inside “Uimahallien kÑyntitapahtumatv2.csv” -file, it seems the text-encoding on the file is off as I'm unable to find the appropriate encoding the display the Finnish letters properly even on my own machine. The file contains the date, time, location and entries product sold.

### Traffic Accidents in Helsinki:

<https://www.avoindata.fi/data/en/dataset/liikenneonnettomuudet-helsingissa>

Unfortunately, the English page for this data doesn't contain the appropriate explanation for the data and it's abbreviations. The information can be found in Finnish on the on the Finnish page( <https://www.avoindata.fi/data/fi/dataset/liikenneonnettomuudet-helsingissa>)

The dataset is a single csv-file hki\_liikenneonnettomuudet.csv, which contains accident category abbreviation, north and east coordinates (which are unused for this assessment), accident severity number and year.

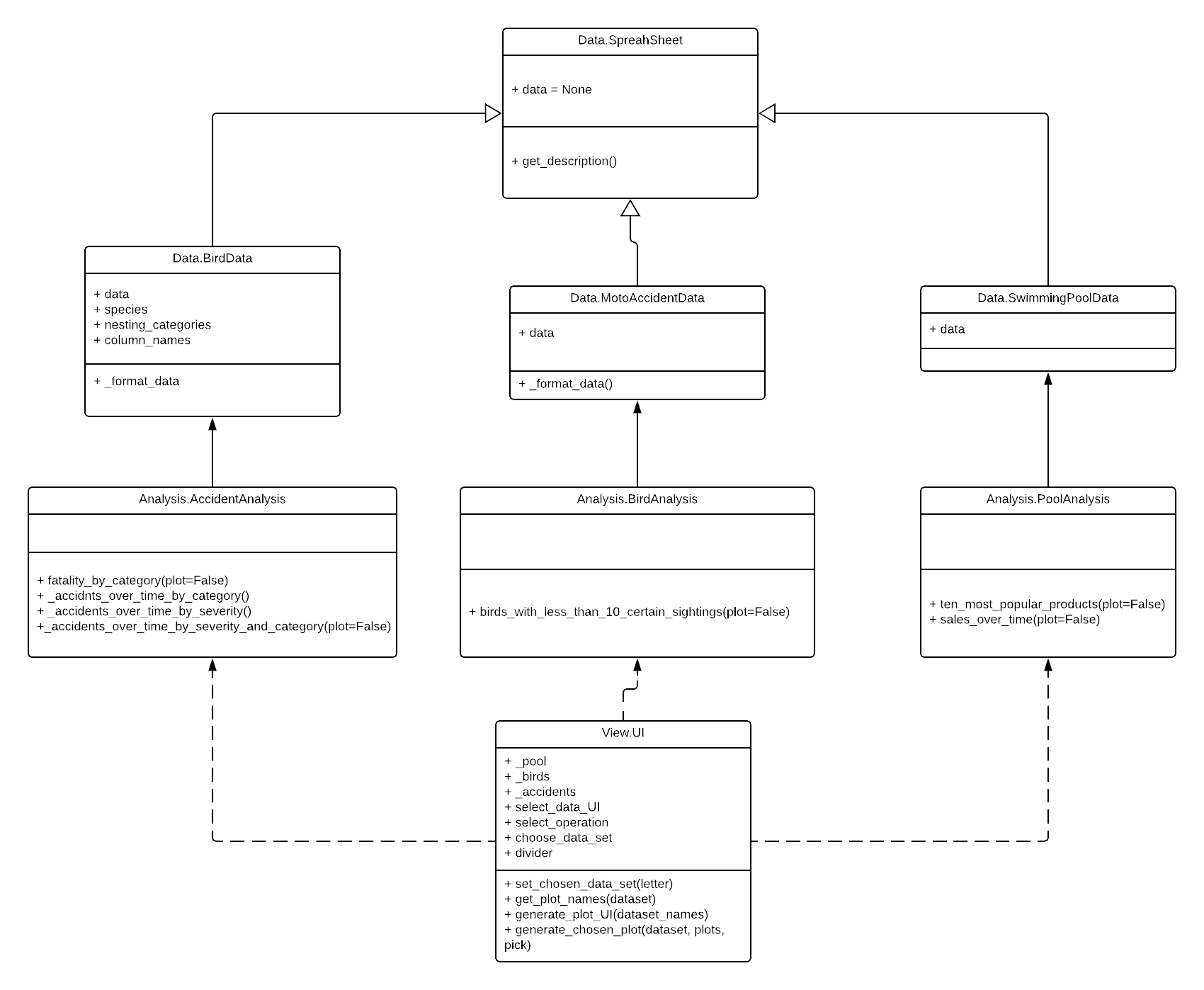
## Demonstrator capabilities

The program has a simple user interface implemented using the click library. From the command line interface, one can choose a dataset they are interested in, choose to either receive basic statistics about the data or to go to another menu to choose a premade plotting option. Ones a plotting option is chosen python will open a new window containing the chart chosen.

The application is fairly inflexible about the source data and requires an additional data and analysis class in addition to a change in the UI to implement a new dataset. But adding new analysis methods to an existing dataset will be rendered in the user interface without requiring any additional code in the View-module.

The plots themselves are fairly simplistic creations as all the datasets are categorical and would require a lot of data massaging to create complex graphs.

## UML



## 

## Examples of output

**>>> import pandas as pd  
>>> import matplotlib.pyplot as plt  
>>> from Analysis import PoolAnalysis, AccidentAnalysis, BirdAnalysis  
>>> pools, accidents, birds = PoolAnalysis(), AccidentAnalysis(), BirdAnalysis()**

**>>> pools.get\_description()**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Date | Time | Location | Product |
| count | 866729 | 866729 | 866729 | 866729 |
| unique | 362 | 53174 | 5 | 53 |
| top | 17.3.2014 | 6:00:45 | Myyrmäki | Uinti Myyrmäki 1x |
| freq | 4903 | 273 | 360520 | 109944 |

**>>> birds.get\_description()**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Species | Nesting category 1974-79 | Nesting category 1986-89 | Nesting categories combined |
| count | 316265 | 316265 | 316265 | 316265 |
| unique | 238 | 5 | 5 | 4 |
| top | Willow Warbler | Certain Nesting | Certain Nesting | Certain Nesting |
| freq | 3703 | 116046 | 109379 | 164594 |

**>>> pools.ten\_most\_popular\_products(plot=True)**

![](data:image/png;base64;base64,)

**>>> accidents.fatality\_by\_category(plot=True)**

![](data:image/png;base64;base64,)

**>>> accidents.accidents\_over\_time\_by\_severity\_and\_category(plot=True)**

![](data:image/png;base64;base64,)

**>>> birds.birds\_with\_less\_than\_10\_certain\_sightings(plot=True)**

![](data:image/png;base64;base64,)