

Project Title

Interim Report

DTXXX

BSc in XXX XXX

**Student Name**

**Student Number**

**Supervisor**

School of Computer Science

Technological University, Dublin

**Date**

Abstract

Declaration

I hereby declare that the work described in this dissertation is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Name

Date

Acknowledgements

Body text

Table of Contents

[1. Introduction 7](#_Toc21978017)

[1.1. Project Background 7](#_Toc21978018)

[1.2. Project Description 7](#_Toc21978019)

[1.3. Project Aims and Objectives 7](#_Toc21978020)

[1.4. Project Scope 7](#_Toc21978021)

[1.5. Thesis Roadmap 7](#_Toc21978022)

[2. Literature Review 8](#_Toc21978023)

[2.1. Introduction 8](#_Toc21978024)

[2.2. Alternative Existing Solutions to Your Problem 8](#_Toc21978025)

[2.3. Technologies you’ve researched 8](#_Toc21978026)

[2.4. Other Research you’ve done 8](#_Toc21978027)

[2.5. Existing Final Year Projects 8](#_Toc21978028)

[2.6. Conclusions 8](#_Toc21978029)

[3. Prototype Design 9](#_Toc21978030)

[3.1 Introduction 9](#_Toc21978031)

[3.2. Software Methodology 9](#_Toc21978032)

[3.3. Overview of System 9](#_Toc21978033)

[3.4. Front-End 9](#_Toc21978034)

[3.5. Middle-Tier 9](#_Toc21978035)

[3.6. Back-End 9](#_Toc21978036)

[3.7. Conclusions 9](#_Toc21978037)

[4. Prototype Development 10](#_Toc21978038)

[4.1. Introduction 10](#_Toc21978039)

[4.2. Prototype Development 10](#_Toc21978040)

[4.3. Front-End 10](#_Toc21978041)

[4.4. Middle-Tier 10](#_Toc21978042)

[4.5. Back-End 10](#_Toc21978043)

[4.6. Conclusions 10](#_Toc21978044)

[5. Testing and Evaluation 11](#_Toc21978045)

[5.1. Introduction 11](#_Toc21978046)

[5.2. Plan for Testing 11](#_Toc21978047)

[5.3. Plan for Evaluation 11](#_Toc21978048)

[5.4. Conclusions 11](#_Toc21978049)

[6. Issues and Future Work 12](#_Toc21978050)

[6.1. Introduction 12](#_Toc21978051)

[6.2. Issues and Risks 12](#_Toc21978052)

[6.3. Plans and Future Work 12](#_Toc21978053)

[6.3.1. GANTT Chart 12](#_Toc21978054)

[Bibliography 13](#_Toc21978055)

# 1. Introduction

**As least 2 pages, but as many as you like**

## Project Background

Some background and literature, start with an interesting fact or a newspaper item

In todays word of constant data sharing and large open source datasets, data exploration and exploration tools are extremly useful tools used to help organization understand their data. Most data is stored in databases and usually exported out as CSV files. As you can imagine reading hundeds of thousands of lines of text is not the best way to go about anaylizing data. For one, humans are prone to make little mistakes and would struggle to keep track of all the data. This is where data visualization can become very useful. Data visualization is taking massive amounts of data and displaying in them in a format that is easy to visualize and understand like maps and graphs.

Data visualization can help us better understand trends and allows us to explorer the data more easily to find causes of certain trends. This is a huge advantage for criminal data. There is an anubundance of datasets on criminal activity from around the world online and free to the public. Visualizing this data can help us better understand the data and explortation tools will help us find potienal causes to crime spikes and where and when they might happen.

## Project Description

An overview of the project

Include a diagram

## Project Aims and Objectives

Overall aim and some milestones along the way to achieve the aim

* Project setup (django, webpack, npm, bootstrap, babel, database)
* collect data (weather, deprevation index, census boundries, population per boundries, crime data)
* normalization of data
* Dataset validation, clean up and migrate to database
* Adding in census borders
* Create REST API to query crime dataset in database
* Create filters for data and display them on map
* add Extra details when a crime marker is clicked on
* create clusters for performance
* Create graphs visulizations
* Create Filters for graphs
* Create functionality to comapre two filter options and visualize them with graphs

## Project Scope

Project scope, what the project isn’t about

## Thesis Roadmap

One sentence explaining what each of the following chapters is about.

# 2. Literature Review

**As least 4 pages, but as many as you like**

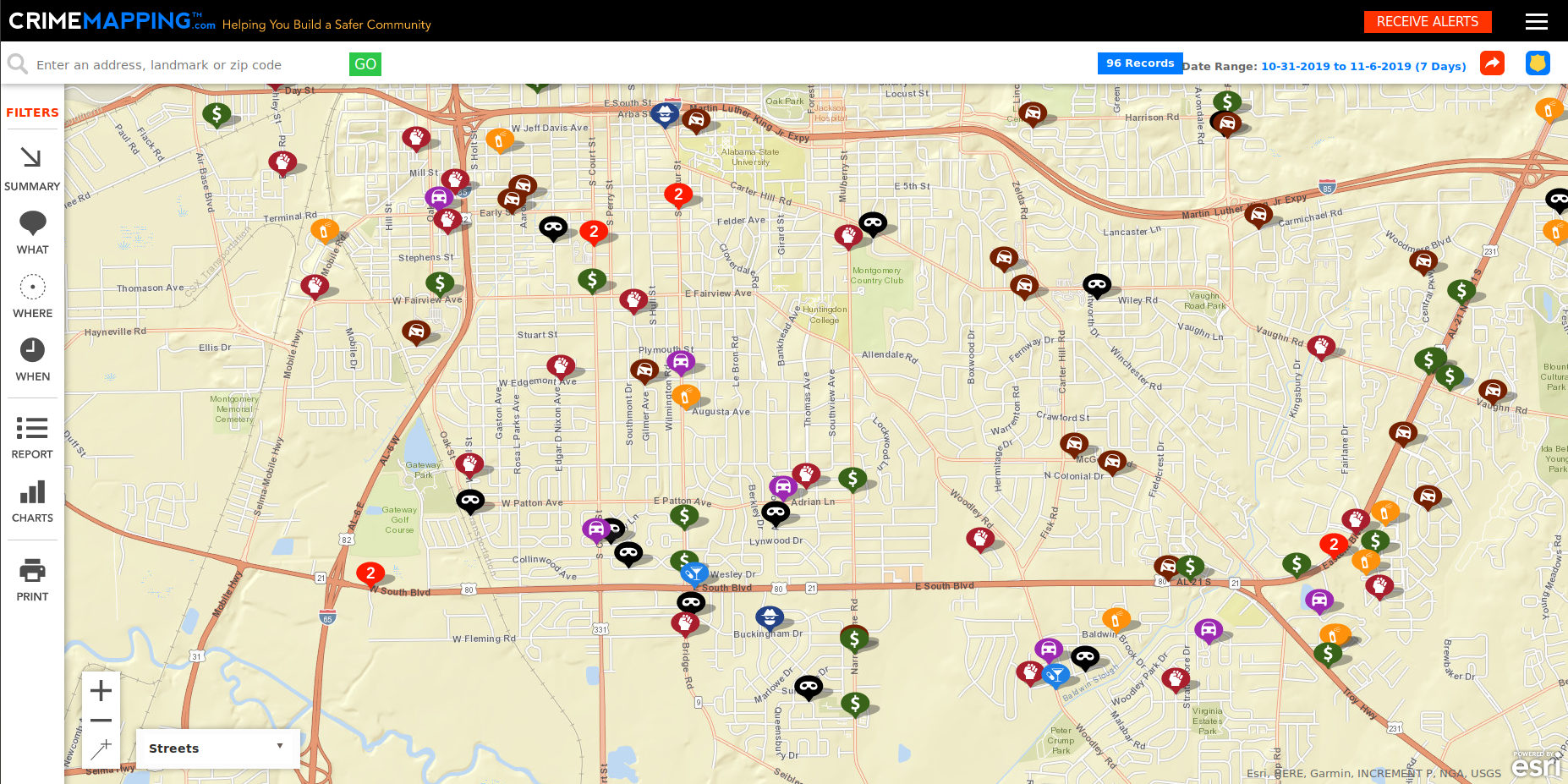
## 2.1. Introduction

In this chapter …

## 2.2. Alternative Existing Solutions to Your Problem

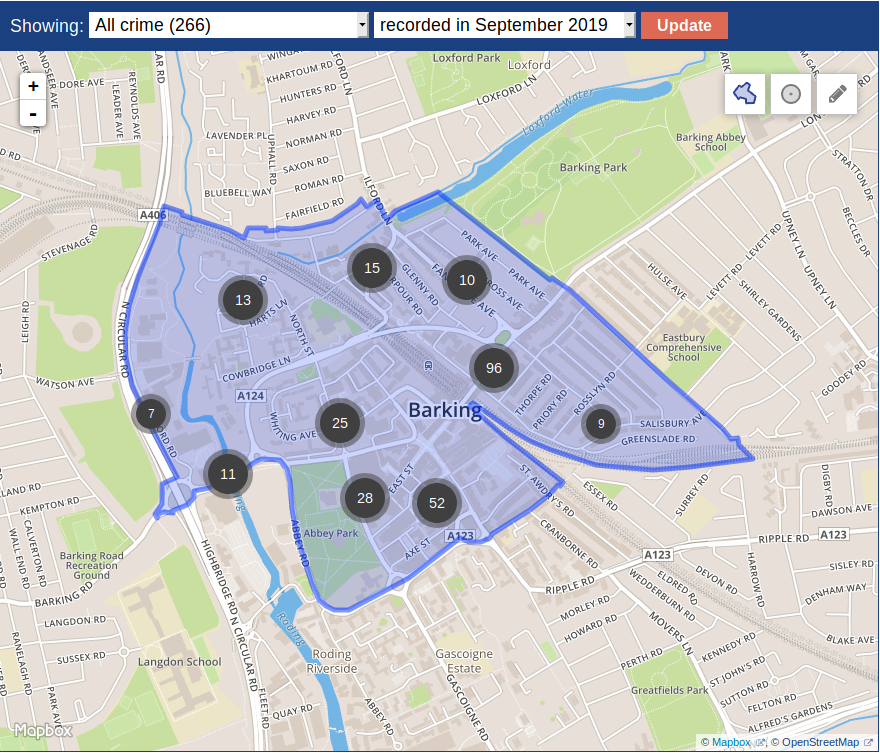
### 2.2.1 Crime Mapping

Crime mapping is a web application that visulaizes all crimes in america. It also provides the user with options to filter for specifc crimes in specific locations [13]. Crime mapping has great visualizations, each crime type has its own map marker and there are options to view the selected crimes in the form of a chart or report. The web application is somewhat scalable. It works well with tablets if they are in landscape mode but some of the filter options are not within the viewport of phone screen. The map performes well on mobile devices and computers. I think is is largerly to do with how it displays crimes. You can only view one thousand crimes at a time on a map. While this will mostly likely increse performance it also restricts the user. This is something to consider when developing my application and might be a comprise I will have to make since the dataset I have chosen has over five million rows of data. Crime Mapping provides the uses with lots of filter options to search for crimes in in your area by a certain date. This is perfect for providing criminal information to the public but it doesn’t have any filters to explore the data. I think it would be usedul to add filters for weather or population in a area to see if these factors have an impact on crime. Below is an image of crimemapping.



### 2.2.2 Police.uk Crime Map

Police.uk. Crime map is another great crime visualization web application. The web application displays all crimes in a spefic month in a niegbourhood. One great feature I noticed from this web application straight away is how it clusters crimes together. This is a huge improvement to the previous crime mapping web applications we looked at (Crime Mapping). This allows the web appliction to display potionally thousnds of crimes on a map without sacraficing performance. However the map doesn’t proved any filter options besides month and nighbourhood since it is mainly used for public information and not data exploration. The neighbour hood zones are a nice feature which allow users to view crimes in a certain area [14]. Below is an image of the web application.



## 2.3. Technologies you’ve researched

### 2.3.1 Web Frameworks

The project will be deployed as a web application and to avoid developing commonaly used features like database abstration models, user autherization and routing I have decided to use a webframe work. There are many webframe works available. Below I will discuss the different types and their advantages and disadvantages.

The first webframe work I am going to look at is Django. Django is a high level web framework built with python. It has a alot of built in features that it comes with out of the box. These features are bundled into componets called apps. Apps can be added to your django project and they can automate a lot of common tasks involved in web development. Tasks such as user authorization and authetencation and much more. Django claims to be extremly fast, scalable and secure1. Companies like instagram, youtube, spotify and bitbucket use django1.

**Pros**

* **Models**: Django has an inbuilt feature where you can define models in a python class. Each field in the model can be converted to a database table. This allows you to define your relational database as a class. You can migrate your models and django will setup a database with models you defined with a database of your choosing. The most powerful and useful feature about django models are once you migrate your models django will provide you with API to query your database through python. This is particually useful when combined with views which are discuesed below 2.
* **Urls**: Another useful feature of django is its url dispatcher. Django alows you to define your own custom urls and map them to a view which will handle the request 3..
* **Views**: Views are how djangos handles requests from the user. When a user goes to a certian url you defined in your url dispatcher it will be mapped to a view. The view will process the request and return a respone e.g. a html page. Views are a big selling point for my project when combined with django model API it will allow me to make request to my database to retrieve data to be visualized 4.
* **Uses Python**: Django uses python which I am very familuar with and has a lot of great libraries.
* **Security**: Django includes a securty layer that helps prevent Cross-site request forgery and sql ingections 5.
* **Admin page**: Django provides you with a built in admin page that allows you to manage your models, users and groups.
* **GeoDjango**: Geodjango is a feature that comes with django that help you make spacial queries to a database. You can define spacial datatypes like multi-polygons and provides you with a model API to make spacial queries e.g. return the country the point(x,y) is in.

**Cons**

* **Everything include**: Django has a lot of features include by default which depending on your application might be overkill and proved unnessarcy overhead and bloat ware.

Flask is a web framework built with python. It is commonly called a micro framework because unlike django it doesn’t come with user autherization or a database abstraction model. However these features can be added to flask using extension. Extension are components at automate alot of common task in web development like user autherization. This way you can build a web application with just the functionality and not have any unwanted modules and bloatware 6.

**Pros**

* **Extensions**: Extension allow you to create a minimal app with just the components you need.
* **Python**: Uses python which has a lot of great libriaries

**Cons**

* **Documentation**: Flasks documention isnt to the same standard as djangos’
* **Project Structure**: While flask is a great and compact webframe work it doesn’t provide you with the same structure as django. Django provides you with a clean structured folder structure when you create a new project.

Ruby on rails is a web framework like flask and django. It automates a lot of the common task in web development and is built on the model view controller MVC philosophy. It uses ruby has its main programming language. It is classifified as an optinionated framework meaning you are encourged to develop your application in way they want to7.

**Pros**

* **Clean code**: Ruby enforces the MVC design pattern and forces you to write cleaner more maintainable code.
* **Community**: Ruby on rails has a great community behind it with lots of posts on stackoverflow.

**Cons**

* **Not flexible**: Ruby on rails is not as flexibale as Flask you can’t easily add extensions to it and it doesn’t have a inbuilt spacial component for dealing with spacial data like django does with geodjango.

After a lot of consideration I have decided django would be the best fit for this project. It provides a lot of commonly need tasks built in and an database abstraction model to easly perform queries to a database and the main reason I have chosen django is because of its geodjango extension app which will allow me to and geo spatial data.

### 2.3.2 Databases

For this application I will be dealing with a lot of large data sets. The criminal records data set just the city of chicago is 2GB. Will this isn’t too much if I were to added in more cities and towns this dataset could be large very quickly. One of the solutions to dealing with the larges datasets is normalizating the data (only if using a relational database). Another big aspect of my application is going to be performace. I want the user to be able to filter crimes and display them on a map as quickly as possible. Taking this into consideration I would need a database with fast read speeds. Another big part of this application will be spatial data like border polygons. I will need a database that can handle these datatypes. In summary the main attributes I am looking for in a database is read speed, scalibilty and can handle spatial data. I have to consider two types of databases SQL and no-SQL. The advantages of a no-SQL database is that it is more flexiable and you can dont have to spend too much time creating relational schemes. Although someone people may see this as a disadvantage depending on the project. No sql also can store data as a json object making it easier to parse and send out in a reqiest response. SQL on the other hand scales better than no-SQL because common values can be stored ina table and referenced with a key. SQL databases generally provide more flexablity for querying data. However this deponds on how you design your scheme. I am leaning towards towards a SQL database because it will provide a more flexible querys, scales well and as a added bonus I am more familar with SQL database. However below I will explore possible options.

The first database I looked at was postgres. Postgresql is a retional open source database it has been ACID complient since 2011. Postgresql is highly extensionable and has a big communtity behind it and a lot of documentation 8. Postgresql caught my eye because of its PostGIS extension. This extension alows you to make spatial querys, for example is point(x,y) in ireland. This could be very useful for my project if I only want to display crimes in a certain county for town. Another big advantage of postgis is that it has spatial datatypes like polygon and multipolygon which can be used to store information about borders which can then be visualized on a map.

MongoDB is also a possible option for this project. It is a no-SQL database and is used by many large companies such as facebook, ebay, google, sap and many more 9. The advantages of mongodb is that it is scalable, what this mean in my situation is that data can be stored in documents that are simular to json. When I want to add a new information that might have a new column for two I can just added it to the database with little to no hassle on the other hand in a SQL database I would have to alter the database table and possibily perform normalization to make sure the database is up to industry standard. This would cost extra development time. Another advantage of mongdo is that because it doesn’t have relations there is no need to developed a relational database model that I would most likely have to reference every no and then to make sure I am querying the right data and depending on how I design it I may need to re-design the model if I want to remove or add new columns. This is not a huge issue but something to consider as it could save me development time. In terms of query performance mongdb would be faster than a SQL database the downside is that if you want to perform more complex queries that need very specific data then noSQL might not be the best option.

In this project I am leaning towards mongodb because it is easy to set up and scales well with newly added data it is also not restricted to a relation schema. The one downside to mongodb is that it doesn’t have any fully featured extensions for handling geo spatial data like multipolygons for borders. While this data can still easily be stored in mongodb, postgresql comes with an extension called postgis which provides custom datatypes another benifit to postgres is that it works well withs geodjangos model api. This means I can easly perform geo spatial queries from django to a postgres database with the postgis extension. For this reason I have decided to choose postgresql for this project.

## 2.4. Other Research you’ve done

### 2.4.2 Datasets

A major factor in this project is going to be what datasets I choose. The datasets need to span span over a decent peroid of time be accutarate and have a lot of information. One of the first places I looked was the UK. The UK has a good dataset with lost of inforamation as well as longitude and latitude positions and a small description of the crime. The UK also has census data which can be used to filter crimes in areas that have a low deprevation index. A deprevation index is a measurement from one to ten of how aflunent or disadvanteged an area is. This is based of census data which includes poverty, population, type of employment/skill and a few others. This can be used to better visualize criminal data in certain areas.

Another possible candinate for datasets is the city of Chicago in America. The reason I am considering this city as a possiblilty is because it has a detailed dataset on crimes going back to 2001 census data that calculates deprevation index of areas and chicago as lot of dataset relation to public facities like transport, education, enviroment and many more 10. In terms of datasets relating to crimes and census data Chicago and the UK are that different, however Chicago as a lot of other useful datasets that could become useful for visualizing criminal data after more research about what causes crime, which will be discussed below.

Orginally I conisdered ireland and a possiblity however while ireland has excellent census data its dataset on crimes is not very accurate and reports crime by garda divions which isn’t very useful for visualizations. It is also marked as “under reservation” by the cental statistics office, due to the data not beign accurate 11.

I have chosen to use Chicago as the focus point for this application

After a quick anyalsis of the data I found there was very little missing values except for the comunity areas, where 8% of the values are missing. Community areas are areas in chicago that are used for gathering census data. Luckily using geodjango I can query the database of comunity area borders and using the longitude an latitude of the crimes and update the missing values.



### 2.4.3 Filter Options

The main part of this web application is going to be the the filter options. There are a lot of factors that could potienally cause crime. I will go through all the filter options I have chosen and why I have chosen them.

Deprevation Index

Community Area

Weather Type

Weather Degrees

crime type

date and time

location (long & lat)

## 2.5. Existing Final Year Projects

### 2.5.1. Deividas Savickas - Crime Explorer

Crime explorer is a very simular in concept to my idea. He created a web application that visualizes criminal data on map and on graphs. The application displays crimes on a map with options to filter the crimes. The options are limited to date (by month), area and crime type. This is perfect for just displaying information to the public. Deividas Savickas gave me a lot of insight on all the possible challenges I could potienally face in the development of this project. One of the issues he faced was the performance of the map when loading too many crime markers. He solved this by using marker clusters. Marker clusters are markers that are close together and grouped into clusters, so 5000 crime markers could potienally be reduced around to 10 clusters. The closer you zoon in the clusters will split and break up until evenually there is just a crime marker left which will be displayed. I am going to be using this technique in my web application because it will provided a sigificant performace boost. Another issue Deividas encoutered was performance issues with loading boundries onto the map. Boundries are polygons that are stored as an array of points. Some boundiers can have up 5000+ points and when you combine that with over 100 different types of bounders all those polygons can have a huge affect on performance. He solved this problem by removing some of the points in each boundrie. This increased performance but decreased the accuracy of the bounderies.

Crime explorer is an excellent web mapping and visualization tool. The main takeaways I got from his project was that when dealing with large datasets performace is going to be challenge and some comprises will have to be made to increase performance and decrease the acuracy of the map.

## 2.6. Conclusions

# 3. Prototype Design

**As least 6 pages, but as many as you like (but lots of diagrams, which count towards the page total).**

## 3.1 Introduction

The prototype will showcase a bit of the main features of the project. The first feature and potienally the most time consuming is the collecting, validating and loading of the data. I am using a lot of different datasets and merging them into one system. A lot of consideration has to be taken into validating the data because the visualization of the data is a big part of the project. If the data is not accurate or up to date the web application will not be useful.

## 3.2. Software Methodology

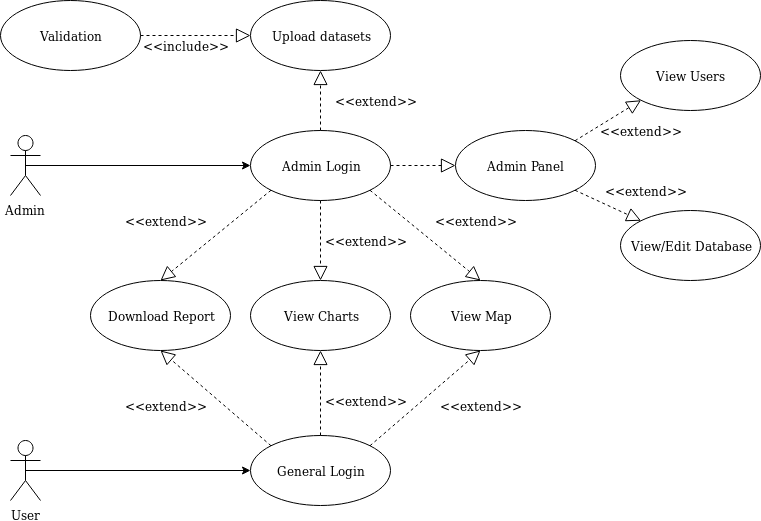
## 3.3. Overview of System

Include a diagram

## 3.4. Front-End

Including screen prototypes and Use Cases

The frontend is going to be an intergral part in the application. There will be two types of users that can use the web application; admin user and general user. The admin user can upload data, edit data, remove data and view logs for the data validation process when a file is uploaded. A general user can view the map and graph data visualizations. Bellow is a use case diagram of the system.



An admin can upload new data to the database. This will allow a none techinical person upload new crimes to the database.

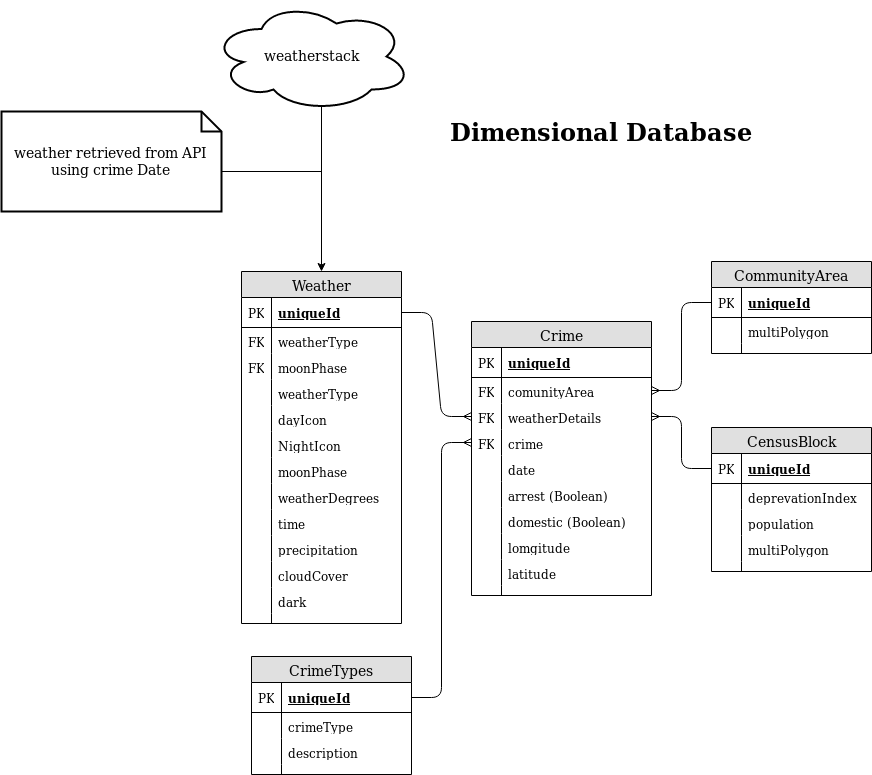
Developing the frontend is going to be another major part of this project.

## 3.5. Middle-Tier

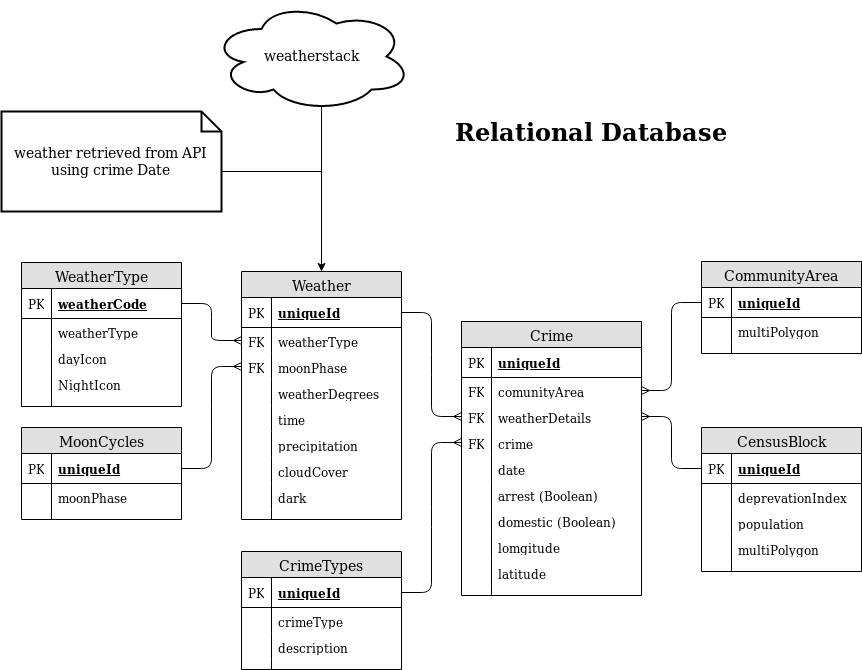
## 3.6. Back-End

Including ERDs, and maybe ISDs

When creating the database, the first step I took was collecting all the data and creating a ERD diagram for the data. First I decided to create a dimensional model for the database scheme. A dimensional model also known as a star shema is a database design used for data warehousing. A central table known as the fact table is used as the main data source and dimensions are added on to it. This makes adding new data easy, all you have to do is create a new dimension and add it to the fact table in constrast to heavy relational designs you might have to redesign the database to add in new data. [16] This is very appealing for my project because based on use feedback I might need to add in new data. Below is a ERD diagram of my first iteration of the database schema.



The dimension model was nearly in the 3rd normal form with the weather table being the exception. In my second iteration I decided to split the weather table up into moonCycles and weather types. While this would impact performance because there will be an extra two tables to query. I decied to do it because it made the shema more readable. Below is my relation design of the database.



This this version of the diagram has an extra two tables. The reason I am choosing this shema as my final design is because the table that defines the weather type and moon cycle ensures all rows in the weather table have predefined weather types and moon cycles. This will make filtering the data easier because all the possible weather types are in the WeatherType table. I can then extract that data and use it in the front end and list all the possible weather types the user can select in the query. This can also be done if I was using the star shema but I would have to check and validate each row and check that it has the correct weather type and stores all the valid weather types somewhere and use them to update the front end filter options. Putting all weather types and moon cycles in their own table makes the development a little easier.

## 3.7. Conclusions

# 4. Prototype Development

**As least 2 pages, but as many as you like (but lots of code samples).**

## 4.1. Introduction

## 4.2. Prototype Development

## 4.3. Front-End

## 4.4. Middle-Tier

## 4.5. Back-End

## 4.6. Conclusions

# 5. Testing and Evaluation

**As least 2 pages, but as many as you like**

## 5.1. Introduction

## 5.2. Plan for Testing

## 5.3. Plan for Evaluation

## 5.4. Conclusions

# 6. Issues and Future Work

**As least 5 pages, but as many as you like**

## 6.1. Introduction

## 6.2. Issues and Risks

## 6.3. Plans and Future Work

### 6.3.1. GANTT Chart

# Bibliography

1 Djangoproject.com. (2019). *The Web framework for perfectionists with deadlines | Django*. [online] Available at: https://www.djangoproject.com/ [Accessed 15 Oct. 2019].

2 Docs.djangoproject.com. (2019). *Models | Django documentation | Django*. [online] Available at: https://docs.djangoproject.com/en/2.2/topics/db/models/ [Accessed 15 Oct. 2019].

3 Docs.djangoproject.com. (2019). *URL dispatcher | Django documentation | Django*. [online] Available at: https://docs.djangoproject.com/en/2.2/topics/http/urls/ [Accessed 15 Oct. 2019].

4 Docs.djangoproject.com. (2019). *Writing views | Django documentation | Django*. [online] Available at: https://docs.djangoproject.com/en/2.2/topics/http/views/ [Accessed 15 Oct. 2019].

5 Docs.djangoproject.com. (2019). *Security in Django | Django documentation | Django*. [online] Available at: https://docs.djangoproject.com/en/2.0/topics/security/ [Accessed 15 Oct. 2019].

6 Flask.palletsprojects.com. (2019). *Welcome to Flask — Flask Documentation (1.1.x)*. [online] Available at: http://flask.palletsprojects.com/en/1.1.x/ [Accessed 16 Oct. 2019].

7 Ruby on Rails Guides. (2019). *Getting Started with Rails — Ruby on Rails Guides*. [online] Available at: https://guides.rubyonrails.org/getting\_started.html [Accessed 16 Oct. 2019].

8 Postgresql.org. (2019). *PostgreSQL: About*. [online] Available at: https://www.postgresql.org/about/ [Accessed 17 Oct. 2019].

9 MongoDB. (2019). *The most popular database for modern apps*. [online] Available at: https://www.mongodb.com/ [Accessed 17 Oct. 2019].

10 Data.cityofchicago.org. (2019). [online] Available at: https://data.cityofchicago.org/ [Accessed 18 Oct. 2019].

11 Cso.ie. (2019). *Crime and Justice - CSO - Central Statistics Office*. [online] Available at: https://www.cso.ie/en/statistics/crimeandjustice/ [Accessed 18 Oct. 2019].

[12] Data.cityofchicago.org. (2019). [online] Available at: https://data.cityofchicago.org/ [Accessed 18 Oct. 2019].

[13] Crimemapping.com. (2019). *CrimeMapping.com - Helping You Build a Safer Community*. [online] Available at: https://www.crimemapping.com [Accessed 30 Oct. 2019].

[14] Police.uk. (2019). *Police.uk*. [online] Available at: https://www.police.uk [Accessed 7 Nov. 2019].

[15] Savickas, D. (2019). Crime Explorer [Accessed 7 Nov. 2019].

[16] Guru99.com. (2019). *What is Dimensional Model in Data Warehouse?*. [online] Available at: https://www.guru99.com/dimensional-model-data-warehouse.html#15 [Accessed 25 Nov. 2019].