

# Crime and Neighbourhoods in Gosport, Hampshire, UK

## IBM Data Science Professional. Course 9 Week 5



### Introduction

Gosport is a town in Hampshire on the south coast of the United Kingdom. At the 2011 Census, its population was 82,622.

It is situated on a peninsula on the western side of Portsmouth Harbour.

Until the last quarter of the 20th century, Gosport was a major naval town associated with the defence and supply infrastructure of Her Majesty's Naval Base (HMNB) Portsmouth.

### Problem

City planners / councillors along with homeowners, residents and families that are interested in relocating to Gosport, will have an interest in the answers to the following questions to help decide which neighbourhood to live in.

1. What neighbourhoods have the highest rates?
2. Is there a relation between crime rate and house price?
3. Are there any areas that are vulnerable to burglaries / vehicle thefts?
4. How well is the area served regard to entertainment venues?
5. How does house price, crimes, venues, population number relate to the neighbourhoods?

## Data

**To investigate these questions, we are going to gather data from various sources**

1. Geolocation data for Neighbourhoods.

In the United Kingdom, neighbourhoods are called Wards. The data will reflect this but be altered to keep a consistent theme. This data was downloaded from the Office of National Statistics (ONS) at their Open Geography Portal (OGP). <https://geoportal.statistics.gov.uk/>

2. Crime data for July 2018 to July 2019.

The crime data is downloaded from Data Police UK. A site for open data about crime and policing in England, Wales and Northern Ireland.

The crime data from this site whilst providing Latitude and Longitude, does not provide Neighbourhood data. Instead the data is split by its LSOA number.

A Lower Layer Super Output Area (LSOA) is a GEOGRAPHIC AREA. Lower Layer Super Output Areas are a geographic hierarchy designed to improve the reporting of small area statistics in England and Wales.

We need to match these LSOA numbers up to the neighbourhood names. <https://data.police.uk/>

3. LSOA data for Neighbourhoods.

These datasets provide the neighbourhood names for each LSOA number. This was downloaded by ONS OGP. <https://geoportal.statistics.gov.uk/>

4. Median House Price for 2018 for Neighbourhoods.

This dataset was filtered by LSOA and then downloaded from the ONS. <https://www.ons.gov.uk/>

5. Population Numbers for each Neighbourhood.

This was filtered and downloaded from the ONS site.

6. Foursquare API will provide venues and location datasets for our chosen locations. <https://foursquare.com/>

## **Summary**

The neighbourhood geolocation data will enable us to properly map crimes by neighbourhood.

The crime data once cleaned and merged with the LSOA data, will enable us to explore and investigate crimes by type and neighbourhood.

The House Price data will allow us to explore any links between house price and crime.

The population numbers per ward will enable us to explore any links between crime and population density.

And finally, the foursquare data will enable us to explore what type, number of places to eat out in the local neighbourhoods.

## Methodology

### Data acquisition and cleaning

#### Basic information about the neighbourhoods and geographical locations.

The neighbourhood data was downloaded as per sources section and consists of 17 different neighbourhoods.

	Neighbourhood	Latitude	Longitude
0	Alverstoke	50.786301	-1.16156
1	Anglesey	50.783199	-1.13577
2	Bridgemary North	50.832001	-1.17331
3	Bridgemary South	50.825401	-1.16649
4	Brockhurst	50.804600	-1.15482
5	Christchurch	50.798500	-1.13104
6	Elson	50.817299	-1.15143
7	Forton	50.804901	-1.14507
8	Grange	50.802200	-1.17246
9	Hardway	50.810200	-1.13804
10	Lee East	50.802700	-1.18437
11	Lee West	50.794498	-1.19333
12	Leesland	50.798401	-1.14155
13	Peel Common	50.819801	-1.17965
14	Privett	50.793098	-1.15631
15	Rowner and Holbrook	50.814400	-1.16343
16	Town	50.792301	-1.12953

Below is a map of Gosport with the different neighbourhoods and boundaries.



**Crime Data, clean and prepare**

Crime data was downloaded as per sources section.

The downloaded data includes Crime ID, Outcomes and other fields that are not required for this report.

Crime ID	Month	Reported by	Falls within	Longitude	Latitude	Location	LSOA code	LSOA name	Crime type	Last outcome category	Context
NaN	2018-07	Hampshire Constabulary	Hampshire Constabulary	-1.149700	51.353134	On or near Southdown Road	E01022553	Basingstoke and Deane 001A	Anti-social behaviour	NaN	NaN
e84dc85ddd0ead4eea7062d...	2018-07	Hampshire Constabulary	Hampshire Constabulary	-1.145593	51.354669	On or near Silverdale Road	E01022553	Basingstoke and Deane 001A	Criminal damage and arson	Investigation complete; no suspect identified	NaN

All the unnecessary columns are stripped out, which leaves the following data.

	Longitude	Latitude	LSOA code	LSOA name	Crime type
0	-1.149700	51.353134	E01022553	Basingstoke and Deane 001A	Anti-social behaviour
1	-1.145593	51.354669	E01022553	Basingstoke and Deane 001A	Criminal damage and arson
2	-1.146700	51.353067	E01022553	Basingstoke and Deane 001A	Violence and sexual offences
3	-1.144834	51.353800	E01022553	Basingstoke and Deane 001A	Other crime
4	-1.155230	51.353821	E01022554	Basingstoke and Deane 001B	Criminal damage and arson

The neighbourhoods are identified by the LSOA code, and the crime data includes all crime in county of Hampshire.

The next step is to remove all crimes not in Gosport and add in the Neighbourhood names to the data frame.

To do this we downloaded the LSOA code and neighbourhood data from ONS, filtered for Gosport.

	LSOA code	Neighbourhood
0	E01022805	Bridgemark South
1	E01022803	Bridgemark South
2	E01022807	Brockhurst
3	E01032882	Grange
4	E01022820	Grange

We then remove all non-Gosport crimes and merged the LSOA data with the crime dataset using neighbourhoods as the reference point.

We then did the same with Median House price from ONS. This was filtered at source and then merged with the crime database.

We repeated this with the Population density from ONS. This was filter at source and then merged with the crime database.

The crime database then consisted of the following fields:

	Longitude	Latitude	LSOA code	LSOA name	Crime type	Neighbourhood	Median House Price	Population
0	-1.181999	50.828340	E01022800	Gosport 001A	Anti-social behaviour	Bridgemark North	220000	4773
1	-1.180570	50.831316	E01022800	Gosport 001A	Burglary	Bridgemark North	220000	4773
2	-1.183082	50.827313	E01022800	Gosport 001A	Criminal damage and arson	Bridgemark North	220000	4773
3	-1.179927	50.825826	E01022800	Gosport 001A	Criminal damage and arson	Bridgemark North	220000	4773
4	-1.180570	50.831316	E01022800	Gosport 001A	Criminal damage and arson	Bridgemark North	220000	4773

## Foursquare Data – Venues

First, was to check that the neighbourhoods as per Foursquare was the same as the neighbourhoods by ONS.

A printout of the neighbourhoods shows that they are using the same 17 names.

Alverstoke  
Anglesey  
Bridgemary North  
Bridgemary South  
Brockhurst  
Christchurch  
Elson  
Forton  
Grange  
Hardway  
Lee East  
Lee West  
Leesland  
Peel Common  
Privett  
Rowner and Holbrook  
Town

The next step was to collect all the venues within 1500 meters of the neighbourhoods.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alverstoke	50.786301	-1.16156	Stokes Bay	50.783198	-1.155257	Beach
1	Alverstoke	50.786301	-1.16156	Pebbles Cafe	50.782148	-1.154836	Café
2	Alverstoke	50.786301	-1.16156	Bayside Cafe	50.786356	-1.169271	Restaurant
3	Alverstoke	50.786301	-1.16156	The Cocked Hat	50.794823	-1.162816	Pub
4	Alverstoke	50.786301	-1.16156	Shell	50.792754	-1.150394	Gas Station

This returns 257 different venues across the neighbourhoods.

We then split the venues into their categories and grouped by the most common venue.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alverstoke	Restaurant	Hotel	Grocery Store	Pizza Place	Beach	Soccer Field	Gas Station	Café	Park	Fried Chicken Joint
1	Anglesey	Supermarket	Harbor / Marina	Fast Food Restaurant	Rugby Stadium	Grocery Store	Gas Station	Pizza Place	History Museum	Café	Boat or Ferry
2	Bridgemary North	Fast Food Restaurant	Hardware Store	Grocery Store	Pub	Supermarket	Bus Stop	Go Kart Track	Coffee Shop	Pet Store	Furniture / Home Store
3	Bridgemary South	Grocery Store	Fast Food Restaurant	Playground	Go Kart Track	Hotel	Skating Rink	Bus Stop	Gym / Fitness Center	Pub	Gastropub
4	Brockhurst	Grocery Store	Pub	Supermarket	Gas Station	Sports Club	Gym / Fitness Center	Hotel	Park	Convenience Store	Pizza Place

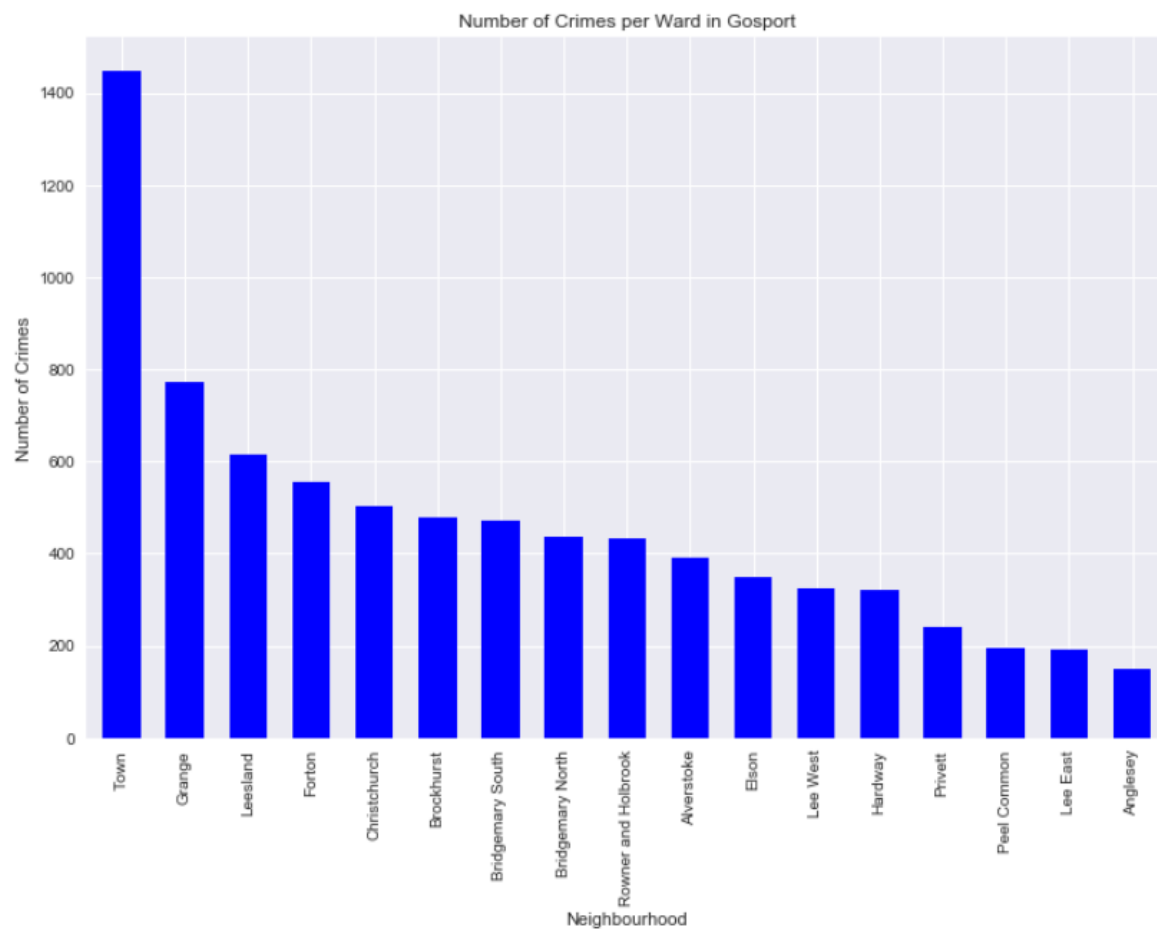
## What is the venue count by Neighbourhood?

	Neighbourhood	Venue Count
0	Alverstoke	14
1	Anglesey	16
2	Bridgemary North	14
3	Bridgemary South	11
4	Brockhurst	20
5	Christchurch	20
6	Elson	11
7	Forton	17
8	Grange	10
9	Hardway	9
10	Lee East	16
11	Lee West	13
12	Leesland	20
13	Peel Common	11
14	Privett	20
15	Rowner and Holbrook	15
16	Town	20



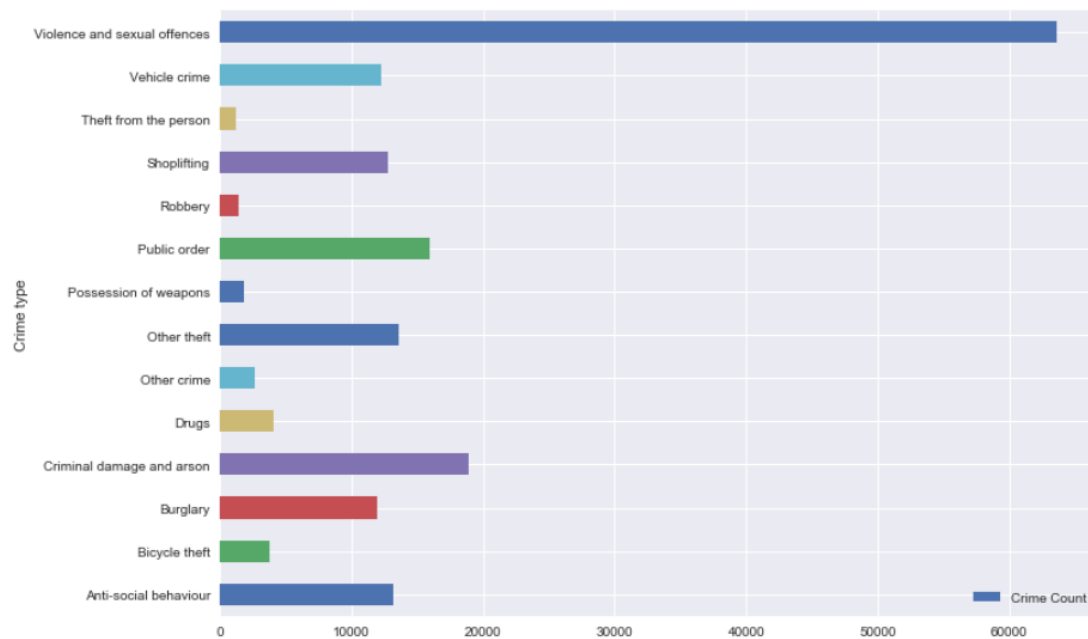
## What's the crime count by Neighbourhood?

	Neighbourhood	Crime Count
0	Alverstone	391
1	Anglesey	151
2	Bridgemary North	438
3	Bridgemary South	473
4	Brockhurst	479
5	Christchurch	503
6	Elson	348
7	Forton	554
8	Grange	771
9	Hardway	320
10	Lee East	190
11	Lee West	325
12	Leesland	615
13	Peel Common	195
14	Privett	240
15	Rowner and Holbrook	433
16	Town	1449

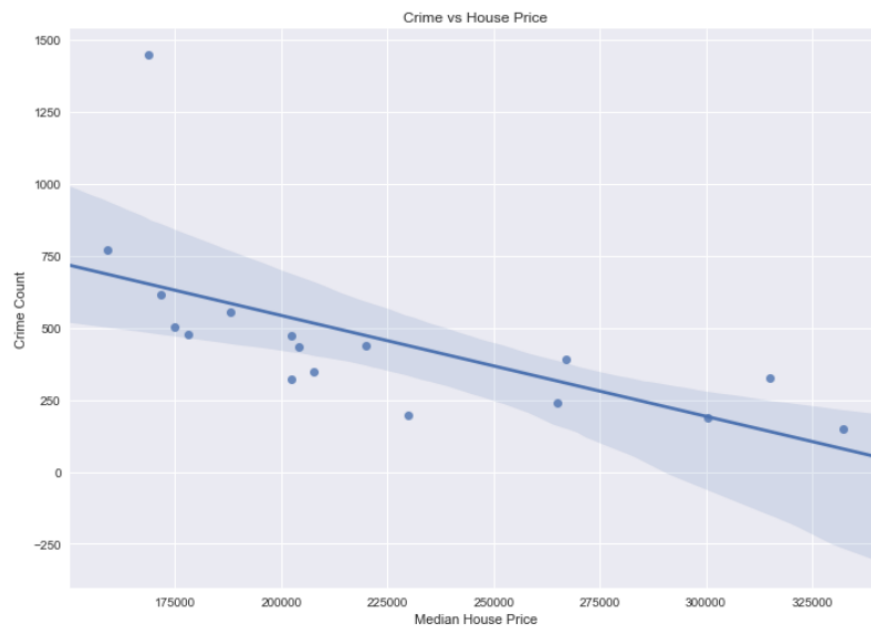


## What type of Crimes happen the most frequently in Gosport?

	Crime type	Crime Count
13	Violence and sexual offences	63563
3	Criminal damage and arson	18940
8	Public order	15909
6	Other theft	13601
0	Anti-social behaviour	13161
10	Shoplifting	12810
12	Vehicle crime	12244
2	Burglary	11940
4	Drugs	4139
1	Bicycle theft	3828
5	Other crime	2708
7	Possession of weapons	1801
9	Robbery	1414
11	Theft from the person	1273

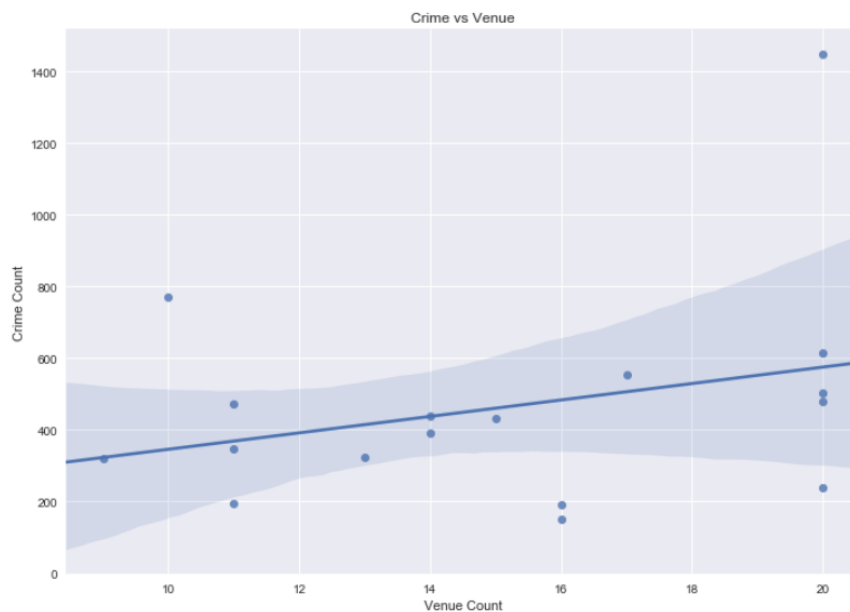


### What is the relationship between Crime and House Price?



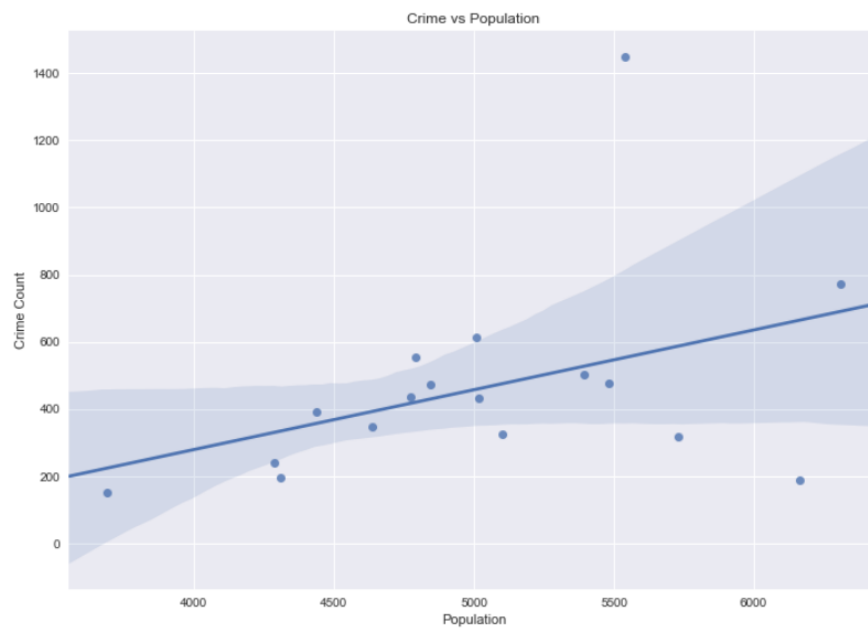
We can see there is a negative correlation between the median house price and the number of crimes.

### What is the relationship between Crime and the Number of Venues?



There is a slight positive correlation between the number of venues and the number of crimes.

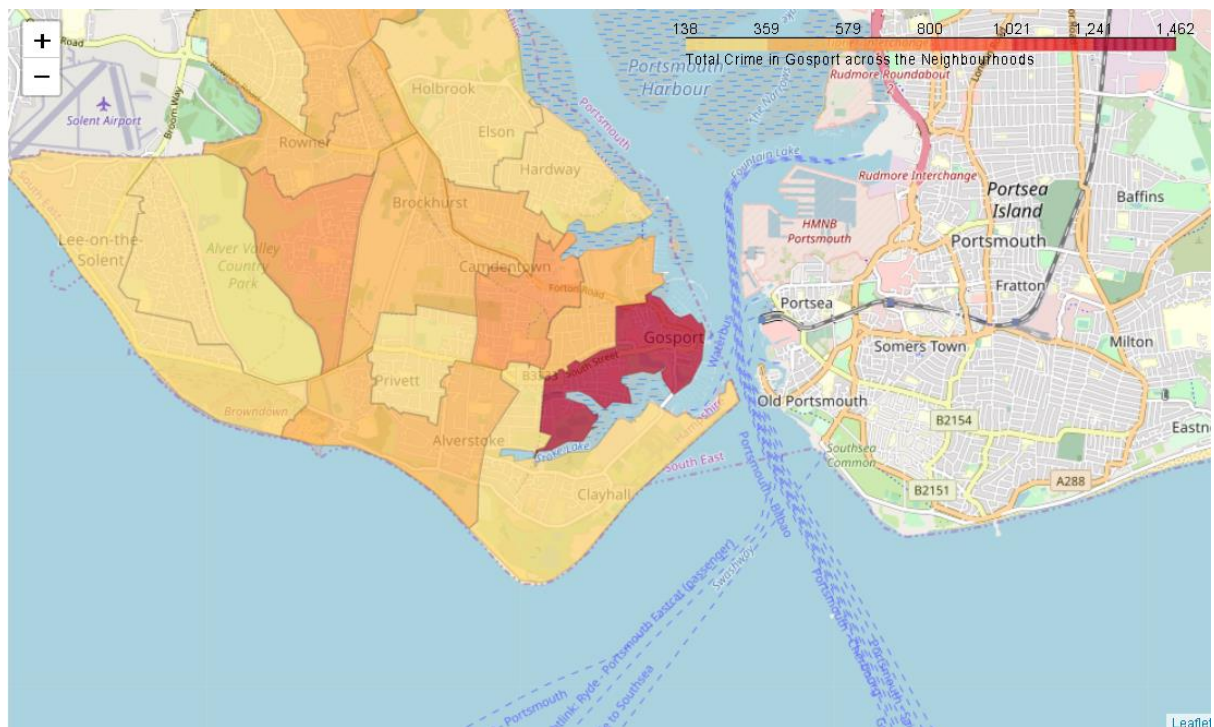
## What is the relationship between Crime and the Population per Neighbourhood?



We can see there is a positive relationship between the population and the number of crimes.

## Total Crime

Certain areas have a total high crime number. Let's see where these neighbourhoods are.



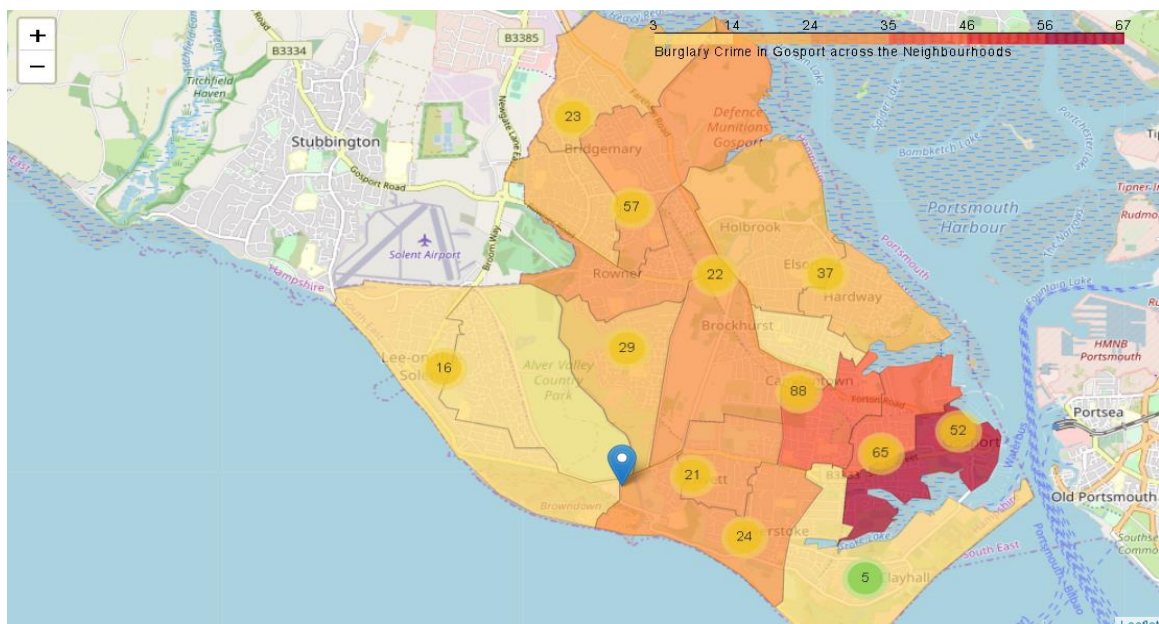
## Burglaries in Gosport

Let's take a closer look at burglaries in Gosport.

We split this data set into its own set on crime type and then group by neighbourhood.

	Neighbourhood	Count
0	Alverstoke	29
1	Anglesey	12
2	Bridgemary North	18
3	Bridgemary South	34
4	Brockhurst	34
5	Christchurch	42
6	Elson	23
7	Forton	13
8	Grange	21
9	Hardway	20
10	Lee East	4
11	Lee West	12
12	Leesland	43
13	Peel Common	17
14	Privett	26
15	Rowner and Holbrook	26
16	Town	66

We then plot the actual latitudes and longitudes of each crime and plot them on the map.



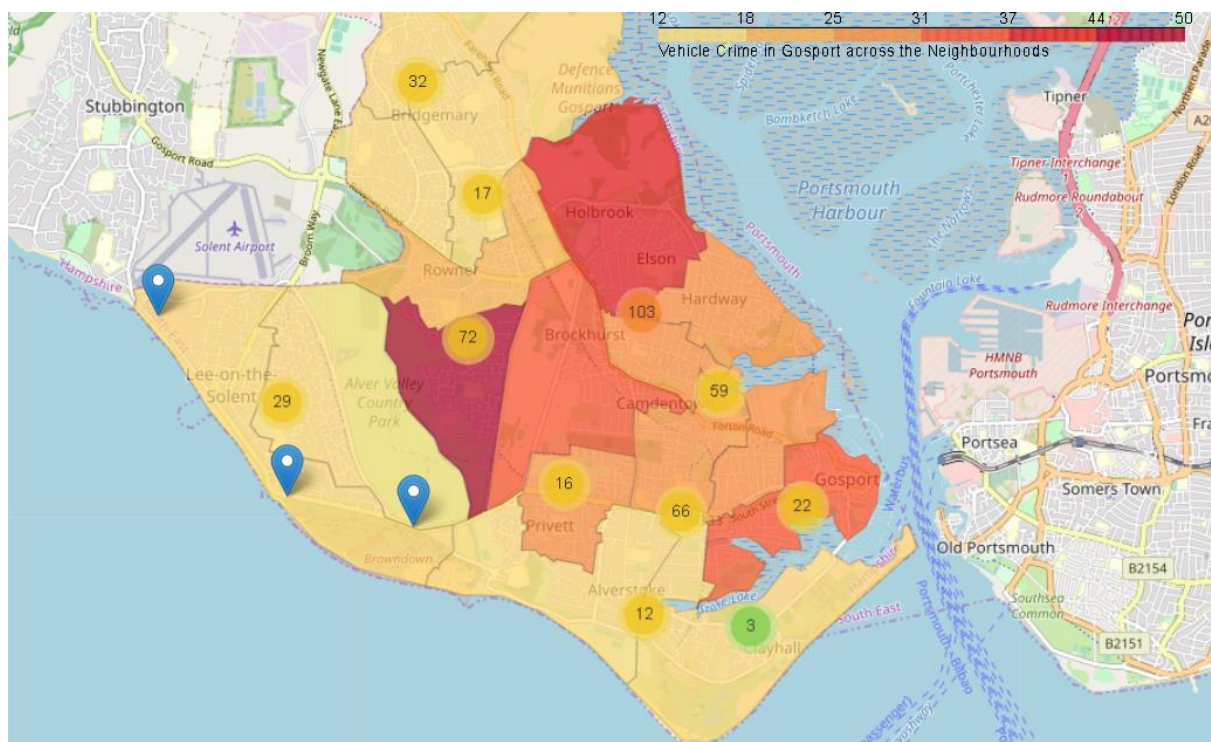
We can see that most Burglaries are committed at the Town and surrounding neighbourhoods.

## Vehicle Crime in Gosport

We split this data set into its own set on crime type and then group by neighbourhood.

	Neighbourhood	Count
0	Alverstoke	16
1	Anglesey	13
2	Bridgemary North	18
3	Bridgemary South	15
4	Brockhurst	34
5	Christchurch	30
6	Elson	39
7	Forton	29
8	Grange	50
9	Hardway	27
10	Lee East	15
11	Lee West	17
12	Leesland	31
13	Peel Common	12
14	Privett	28
15	Rowner and Holbrook	24
16	Town	36

We then plot the actual latitudes and longitudes of each crime and plot them on the map.

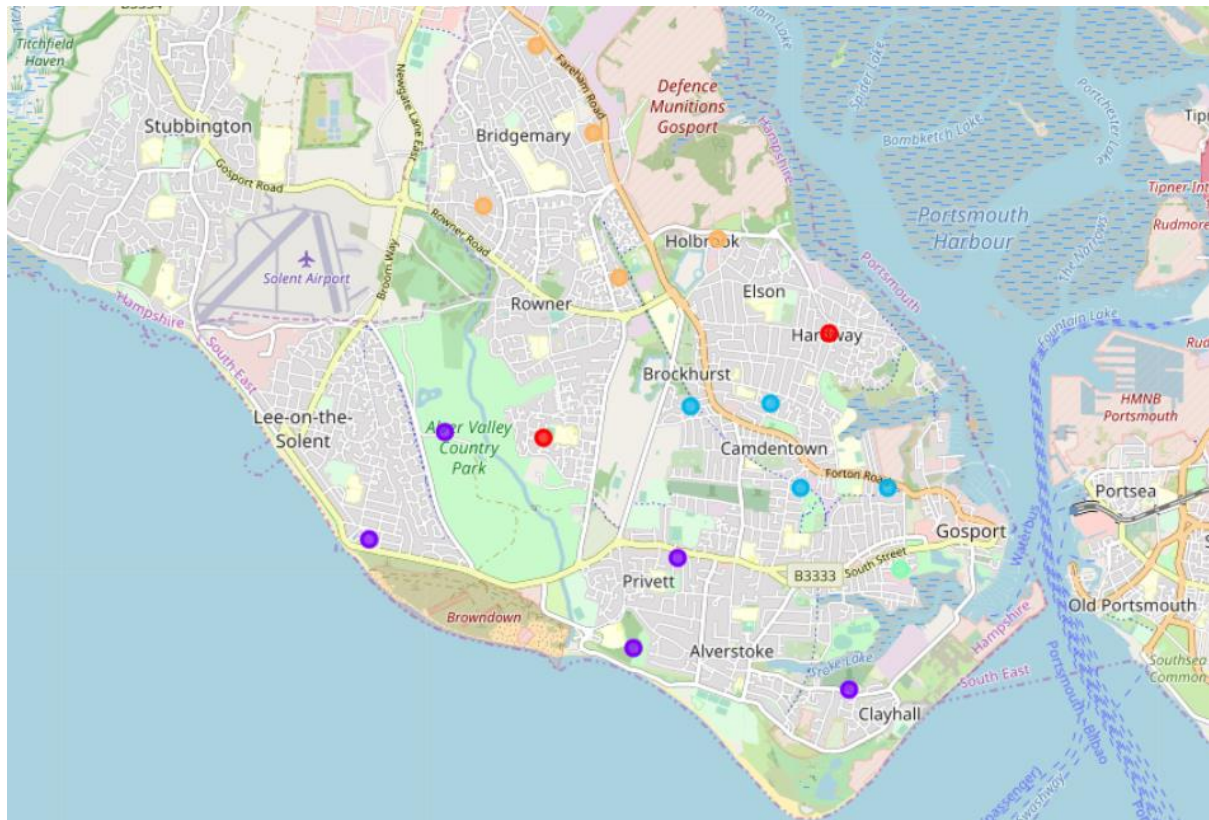




## Cluster Analysis

We then look at the cluster analysis using 5 clusters and the features of Population, Total Crime, Median House Price and Total Venues per Neighbourhood.

We tested the cluster analysis using both raw and scaled features, and in the end settled on using the scaled features due to the cluster analysis would deal with the outlier in the Town features and also that there is not a large change in the range of house price and population whilst there is a large range in the number of crimes.



Before we look at the clusters, let's look at the average, minimum and maximum values of our features.

Crime Min	151
Crime Max	1449
Crime Avg	463
House Price Min	159000
House Price Max	332433
House Price Avg	222718
Venue Count Min	10
Venue Count Max	20
Venue Count Avg	15
Population Min	3688
Population Max	6309
Population Avg	5029

Let's look at the five clusters.

### Cluster 1 – Red

	Neighbourhood	Cluster Labels	Crime Count	Median House Price	Venue Count	Population
8	Grange	0	771	159000	10	6309
9	Hardway	0	320	202250	10	5731

This is a high population with below average house price, low venue count and average to high crime count.

### Cluster 2 – Purple

	Neighbourhood	Cluster Labels	Crime Count	Median House Price	Venue Count	Population
0	Alverstoke	1	391	267000	14	4436
1	Anglesey	1	151	332433	17	3688
10	Lee East	1	190	300500	18	6164
11	Lee West	1	325	315000	14	5102
14	Privett	1	240	265000	20	4287

This is low crime, above average house price, high venue count and a mix of low to high populations areas.

### Cluster 3 – Blue

	Neighbourhood	Cluster Labels	Crime Count	Median House Price	Venue Count	Population
4	Brockhurst	2	479	178000	20	5482
5	Christchurch	2	503	175000	20	5392
7	Forton	2	554	188118	17	4793
12	Leesland	2	615	171568	20	5008

This is a high crime area with cheap house prices and high venues with high population.

### Cluster 4 – Green

	Neighbourhood	Cluster Labels	Crime Count	Median House Price	Venue Count	Population
16	Town	3	1449	168600	20	5538

The town has its' own cluster. It is highly population with maximum venues, extreme crime and cheap housing.



## Cluster 5 – Tan

	Neighbourhood	Cluster Labels	Crime Count	Median House Price	Venue Count	Population
2	Bridgemary North	4	438	220000	13	4773
3	Bridgemary South	4	473	202250	10	4844
6	Elson	4	348	207500	11	4637
13	Peel Common	4	195	230000	11	4307
15	Rowner and Holbrook	4	433	204000	15	5018

This is low to average population, average crime, average house price, and average venue count.

### Discussion

1. What neighbourhoods have the highest rates?

The town are by far has the highest crime rate.

Other neighbourhoods of above average crime are:

Grange  
Brockhurst  
Christchurch  
Forton  
Leesland  
Bridgemary South

2. Is there a relation between crime rate and house price?

There is a negative relationship between house price and crime. The areas with the highest crime have the cheapest house prices.

There is a positive relationship with Venues and Population. The more venue or residents then there tends to be more crime.

3. Are there any areas that are vulnerable to burglaries / vehicle thefts?

Neighbourhoods of note for Burglaries:

	Neighbourhood	Count
16	Town	66
12	Leesland	43
5	Christchurch	42
3	Bridgemary South	34
4	Brockhurst	34
0	Alverstoke	29
14	Privett	26
15	Rowner and Holbrook	26
6	Elson	23
8	Grange	21
9	Hardway	20
2	Bridgemary North	18
13	Peel Common	17
7	Forton	13
1	Anglesey	12
11	Lee West	12
10	Lee East	4

Neighbourhoods of Note for Vehicle Crimes:

	Neighbourhood	Count
8	Grange	50
6	Elson	39
16	Town	36
4	Brockhurst	34
12	Leesland	31
5	Christchurch	30
7	Forton	29
14	Privett	28
9	Hardway	27
15	Rowner and Holbrook	24
2	Bridgemary North	18
11	Lee West	17
0	Alverstoke	16
3	Bridgemary South	15
10	Lee East	15
1	Anglesey	13
13	Peel Common	12

#### 4. How well is the area served regard to entertainment venues?

Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Alverstoke	Grocery Store	Restaurant	Hotel	IT Services	Gas Station
Anglesey	Supermarket	Fast Food Restaurant	Grocery Store	Gas Station	Beach
Bridgemary North	Hardware Store	Fast Food Restaurant	Grocery Store	Pub	Furniture / Home Store
Bridgemary South	Fast Food Restaurant	Grocery Store	Gym / Fitness Center	Go Kart Track	Skating Rink
Brockhurst	Grocery Store	Pub	Gas Station	Supermarket	Convenience Store
Christchurch	Supermarket	Grocery Store	History Museum	Fast Food Restaurant	Convenience Store
Elson	Grocery Store	Pub	Fast Food Restaurant	Gym / Fitness Center	Skating Rink
Forton	Grocery Store	Pub	Gas Station	Convenience Store	Park
Grange	Grocery Store	Supermarket	Soccer Field	Sandwich Place	Pub
Hardway	Pub	Grocery Store	Convenience Store	Harbor / Marina	Supermarket
Lee East	Pub	Café	Sandwich Place	Grocery Store	Supermarket
Lee West	Pub	Café	Grocery Store	Chinese Restaurant	Italian Restaurant
Leesland	Grocery Store	Supermarket	Fast Food Restaurant	Convenience Store	Pizza Place
Peel Common	Grocery Store	Pub	Gym / Fitness Center	Hotel	Fast Food Restaurant
Privett	Grocery Store	Supermarket	Pub	Hotel	Soccer Field
Rowner and Holbrook	Grocery Store	Convenience Store	Supermarket	Gym / Fitness Center	Skating Rink
Town	Supermarket	Pub	Coffee Shop	Fast Food Restaurant	Warehouse Store

As we can see most neighbourhoods have a good selection and variety of venues.

#### 5. How does house price, crimes, venues, population number relate to the neighbourhoods?

The previous analysis shows that there are relationships between crime and house price, venue number and the population number in across the neighbourhoods.

We have used this to cluster neighbourhoods into 5 different clusters with different characteristics.

These can be used to choose to identify neighbourhoods for different purposes. For example, that for a certain house price we could look for neighbourhoods with low crime.

The best neighbourhood in Gosport is the Cluster 2 Purple – which is characterised by low crime, high house prices, good venue count, which also tend to be the neighbourhoods that are the closest to the beach.

### Conclusion

The data has shown that Gosport neighbourhoods have a varying mixture of crime count, population, venue number and median house price.

The data has also show that these features are related to the number of crimes, whether positively or negatively.

And if one wants to choose a low crime neighbourhood, then a high house price, low population, and below average venue number are good indicates of what the crime numbers would be for the neighbourhood.