ANALYZING THE INFLUENCE OF UNEMPLOYMENT ON CRIME RATES IN THE UNITED KINGDOM

Sheffield Hallam University

PROJECT OVERVIEW

EXECUTIVE SUMMARY

This proposal outlines an investigation into the potential nexus between recipients of job seeker allowance and crime incidence within the context of the United Kingdom. The research seeks to yield valuable insights for informing policy formulations and strategic interventions aimed at crime prevention and socio-economic amelioration. Through meticulous analysis of crime rates, unemployment metrics, and demographic variables, the study intends to employ sophisticated analytical techniques to uncover correlations, detect patterns, and identify trends. Rigorous statistical examination and adept econometric modeling will be applied to elucidate the intricate interplay between unemployment dynamics, demographic shifts, and crime rates across diverse geographic regions in the UK. The overarching objective is to contribute to enhanced community well-being and security while furnishing evidence-based guidance for policy decisions.

Sheffield Hallam University

PROJECT OVERVIEW

RESEARCH AIM

To investigate the influence of job seeker allowance (JSA) holders on the crime rate in United Kingdom, providing evidence-based insights for policy and intervention efforts, and enhancing community security and well-being.

RESEARCH OBJECTIVES

1. Investigate the relationship between job seeker allowance holders and crime rates in different regions of United Kingdom.
2. Analyse the correlation between criminal activities and JSA sectors, identifying industries or job sectors with higher incidences of crime and their impact on employment in United Kingdom.
3. Examine the vulnerability of specific demographic groups to job seeker allowances and crime, exploring the intersection of these factors and proposing strategies to overcome this vulnerability.
4. Assess the impact of changes in population demographics, such as migration and shifts in age distribution, on crime rates and employment patterns.
5. Examine JSA distributions in United Kingdom to understand the potential influence of a growing young job seekers on crime rates.

Sheffield Hallam University

BUSINESS QUESTION AND JUSTIFICATION

Business Question	Justification	Dimensions	Metrics
Provide a monthly breakdown of crimes by crime type and by location in UK between 2013 to 2022.	This will provide stakeholders with insights into which location has the highest crime and what type of crime is prevalent within the given period. Which can be used to allocate security resources more effectively, plan infrastructure improvements.	DateDim, CrimeTypeDim, DistrictDim	Crime count
Provide a monthly breakdown of crimes, by location, by job seeker allowance gender in UK between 2013 to 2022.	The analysis aims to reveal links between socioeconomic factors and crime, guiding focused policies, resource distribution, and social support for addressing root causes of criminal activity and promoting a safer, fairer society.	DateDim, GenderDim, DistrictDim	Crime count, JSA count
Provide a monthly breakdown of top three crime types reported in each police force by location and top three job seeker allowance claims by age groups in UK between 2013 to 2022.	This analysis will help to ascertain the primary crime categories in various police jurisdictions, along with age groups receiving the highest job seeker allowance, which could facilitate targeted law enforcement strategies, resource allocation, and social assistance programs to address specific crime challenges and age-related employment needs, contributing to improved community well-being and security	DateDim, CrimeTypeDim, AgeDim, PoliceDepDim	Crime count, JSA count
Provide a monthly breakdown of crime count by region and the corresponding top 3 jobseekers allowance count by age and gender between 2013 to 2022.	This analysis will help to understand potential correlations between crime rates and socioeconomic factors, guiding the formulation of targeted policies and resource allocation, which can aid in developing tailored support systems and interventions to address unemployment-related challenges and foster a more inclusive and secure society.	DateDim, RegionDim, AgeDim, GenderDim	Crime count, JSA count
Provide a month-by-month comparison of top 5 locations in the UK with the highest levels of crime caused by sexual violence, and total job seeker allowance received by people in those locations between 2013 and 2022.	This analysis helps identify potential correlations between spikes in sexual violence and changes in job seeker allowances, enabling targeted interventions, resource allocation, and policy adjustments to address these issues effectively.	DateDim, DistrictDim	Crime count, JSA count

Sheffield Hallam University

DATASET OVERVIEW AND LINK

Crime dataset:

Crime data is taken from the police official website which contains data from 2013 to 2022. The variables of interest for this dataset are Date, Crime type, Crime outcome, Location. These variables have been extracted from the original dataset using R programming. Link to the original dataset is:

https://data.police.uk/data/archive/

Crime ID	Month	Reported by	Falls within	Longitude	Latitude	Location	LSOA code	LSOA name	Crime type	Last outcome category	Context
	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.50993	51.410873	On or near Ludlow Close	E01014399	Bath and North East Somerset 001A	Anti-social behaviour		
	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.513718	51.429407	On or near A4174	E01014399	Bath and North East Somerset 001A	Anti-social behaviour		
	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.511761	51.409366	On or near Caernarvon Close	E01014399	Bath and North East Somerset 001A	Anti-social behaviour		
	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.513718	51.429407	On or near A4174	E01014399	Bath and North East Somerset 001A	Anti-social behaviour		
0437c1dffc09bd01c5f25155ac19ecba4d6571e2afc80aee5611ac795619110b	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.512773	51.411751	On or near Westfield Close	E01014399	Bath and North East Somerset 001A	Burglary		
2fa67d93ccc27fa0583d6d776963b16ead8579db197e6f18ec365cef6f58b736	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.509126	51.416137	On or near St Francis Road	E01014399	Bath and North East Somerset 001A	Burglary		
80cc404d09e2b10e3a3e3b0339afbdb95446bffc63a7bf5b6c7e83c9ef505b41	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.512773	51.411751	On or near Westfield Close	E01014399	Bath and North East Somerset 001A	Burglary		
003778489010ace0465fe263a1cd3221e1e98c7dabc17fd18f60f944543c8fdc	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.509384	51.40959	On or near Barnard Walk	E01014399	Bath and North East Somerset 001A	Criminal damage and arson		
90d416fb81abe3b0b28368eb07817e295b2ee4764e3c760c570826a5541dee79	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.509384	51.40959	On or near Barnard Walk	E01014399	Bath and North East Somerset 001A	Criminal damage and arson		
c7d5a674fbce067977ece3f391265e74329ddcd258ee3e0aabbcee5f87587cbb	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.511927	51.409435	On or near Harlech Close	E01014399	Bath and North East Somerset 001A	Violent crime		
	2012-02	Avon and Somerset Constabulary	Avon and Somerset Constabulary	-2.510427	51.423181	On or near Durley Lane	E01014400	Bath and North East Somerset 001B	Anti-social behaviour		

Fig 1.0: Crime Dataset

Sheffield Hallam University

DATASET OVERVIEW AND LINK

Job Seekers Allowance Data:

Dataset for job seeker allowance has been extracted from the government official website which contains 8 columns for different characteristics of the job seekers. Variables of interest, Date, Age, Gender, Location, has been extracted and modified through R programming to do better investigation with the crime data. Link to the original dataset is:
https://www.nomisweb.co.uk/datasets/ucjsa

	date	Gender	Age	geogcode	value
1	January 2013	Male	Aged 16-17	E06000005	0
2	February 2013	Male	Aged 16-17	E06000005	0
3	March 2013	Male	Aged 16-17	E06000005	0
4	April 2013	Male	Aged 16-17	E06000005	5
5	May 2013	Male	Aged 16-17	E06000005	5
6	June 2013	Male	Aged 16-17	E06000005	5
7	July 2013	Male	Aged 16-17	E06000005	5
8	August 2013	Male	Aged 16-17	E06000005	5
9	September 2013	Male	Aged 16-17	E06000005	0
10	October 2013	Male	Aged 16-17	E06000005	0

Fig 2.0: Crime Dataset

	year	month	Region_code	Region_name	District_code	District_name	Age	Gender	value
1	2013	1	E12000001	North East	E06000001	Hartlepool	16-17	Female	0
2	2013	1	E12000001	North East	E06000001	Hartlepool	16-17	Male	0
3	2013	1	E12000001	North East	E06000001	Hartlepool	18-24	Female	465
4	2013	1	E12000001	North East	E06000001	Hartlepool	18-24	Male	960
5	2013	1	E12000001	North East	E06000001	Hartlepool	25-29	Female	220
6	2013	1	E12000001	North East	E06000001	Hartlepool	25-29	Male	550
7	2013	1	E12000001	North East	E06000001	Hartlepool	30-34	Female	155
8	2013	1	E12000001	North East	E06000001	Hartlepool	30-34	Male	375
9	2013	1	E12000001	North East	E06000001	Hartlepool	35-39	Female	145
10	2013	1	E12000001	North East	E06000001	Hartlepool	35-39	Male	305

Fig 3.0: Job seeker allowance Dataset

EMPLOYMENT PREVIEW BEFORE TRANSFORMATION

DEVELOPMENT METHODOLOGY

The software development methodology used in this project is the Waterfall development methodology.

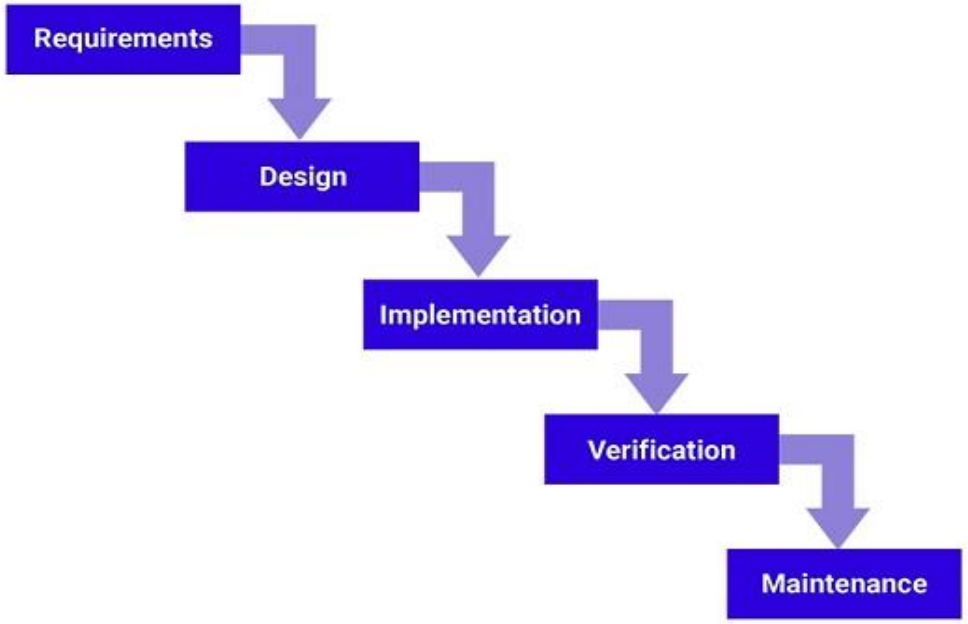


Fig 4.0: Development Methodology

The requirements for this project were gathered from the business case being considered. The team discussed the project plan and began implementation.

Sheffield Hallam University

DEVELOPMENT METHODOLOGY & JUSTIFICATION

The Kimball Bottom-Up approach was used throughout the development process, specifically chosen for its ability to deliver value quickly and iteratively. It ensures data integrity by starting at the lowest level and thoroughly validating and cleansing the data. Additionally, it is scalable, accommodating the integration of additional data sources and expanding the analysis scope.

By starting with the most critical business processes, we were able to develop data marts incrementally, providing stakeholders with quick wins and demonstrating the value of the project. This approach also aligns with the goal of developing a scalable and modular solution that can be easily extended or modified in the future, which is crucial for stakeholders as their business needs evolve.

Overall, the bottom-up approach provides a solid foundation for data-driven decision-making and offers a detailed view of the data for comprehensive analysis. The architectural view of this approach can also be seen in Figure 2 in the previous slide.

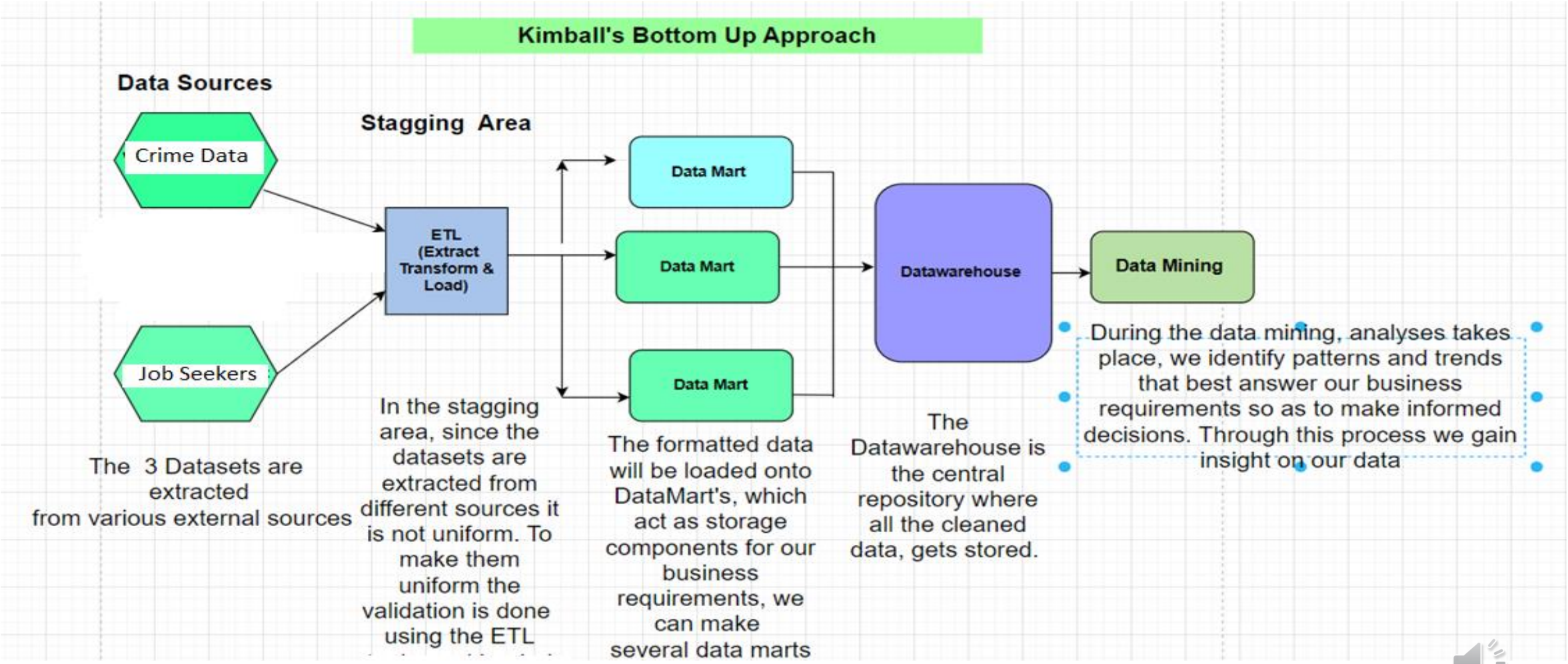


Fig 5.0: Kimball's Bottom-up approach

BDDS TOOLS & DI TECHNIQUES

Sheffield Hallam University

BDDS Tools and DI Techniques

- Apache Hive
- Azure Virtual Machine
- Hortonworks – HDP and HDF
- MS Access
- R
- R Studio
- Tableau

- The entire project development was done on the Azure Virtual Machine environment which provided the flexibility to scale up or down the computing resources needed. It also provided more RAM for handling the large size of data being extracted into R objects for processing.
- R studio served as the staging environment where extraction, cleaning and all validation checks were carried out.
- Apache Hive was connected to Hadoop Distributed File System (HDFS) and used to create tables in which the processed data will be loaded.
- R was connected to Apache Hive using the Database Interface (DBI) and Open Database Connectivity (ODBC) libraries to load the data marts into Hive tables.
- The dimensions and fact tables were replicated in MS Access to generate the star schema. A constellation was used because of the multiple fact tables based on the business questions
- Hive was connected to Tableau via the Hadoop Cloudera server connector for visualisation/Analytical processing.

Sheffield Hallam University

Staging Diagram

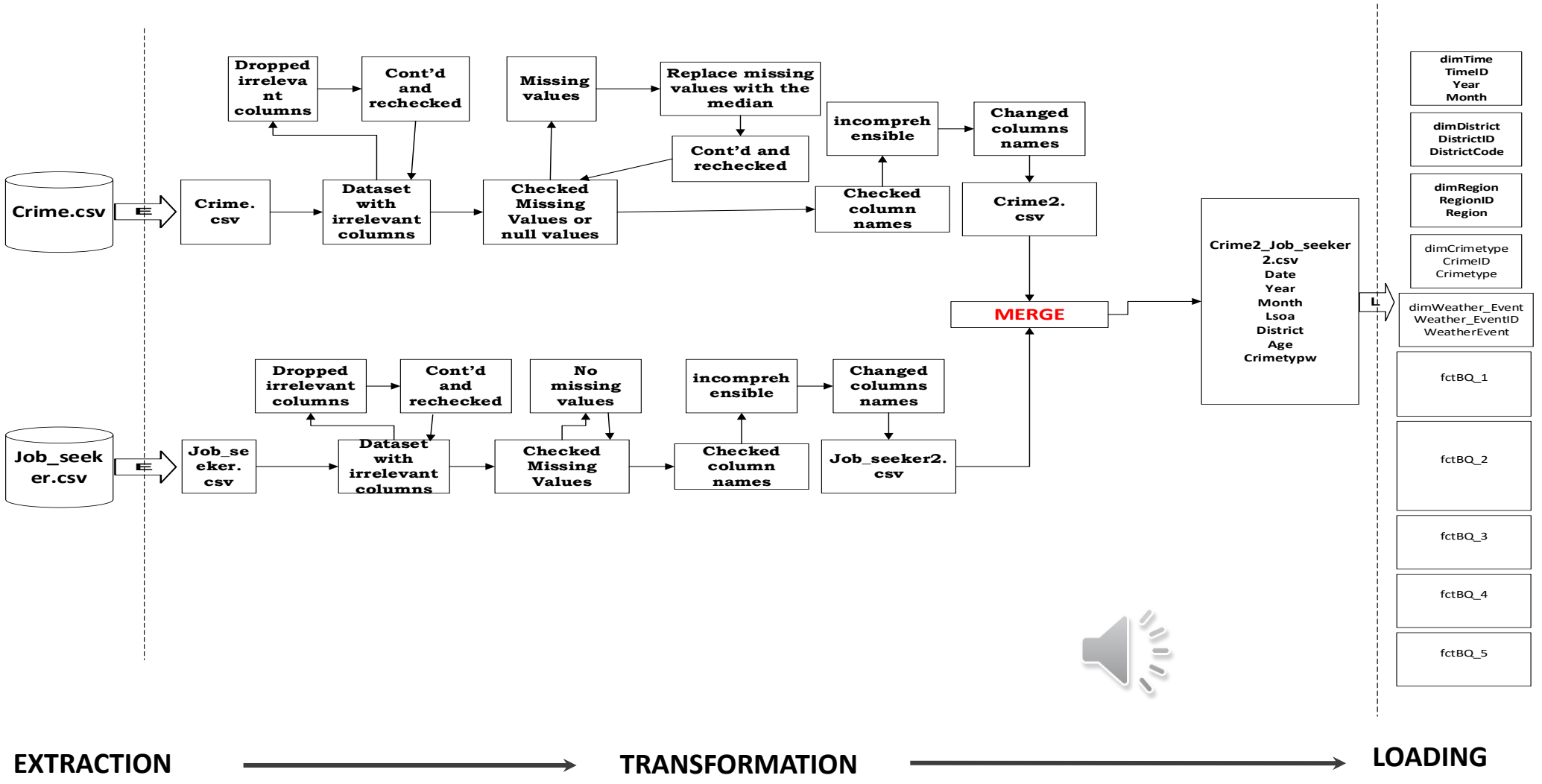


Fig 6.0 Stages involved in data preprocessing leading up to the establishment of a data mart.

Sheffield Hallam University

	date	Gender	Age	geogcode	value
1	January 2013	Male	Aged 16-17	E06000005	0
2	February 2013	Male	Aged 16-17	E06000005	0
3	March 2013	Male	Aged 16-17	E06000005	0
4	April 2013	Male	Aged 16-17	E06000005	5
5	May 2013	Male	Aged 16-17	E06000005	5
6	June 2013	Male	Aged 16-17	E06000005	5
7	July 2013	Male	Aged 16-17	E06000005	5
8	August 2013	Male	Aged 16-17	E06000005	5
9	September 2013	Male	Aged 16-17	E06000005	0
10	October 2013	Male	Aged 16-17	E06000005	0

```
> dataset
                  date Gender        Age     geogcode value
      1:    January 2013   Male Aged 16-17    E06000005     0
      2:   February 2013   Male Aged 16-17    E06000005     0
      3:      March 2013   Male Aged 16-17    E06000005     0
      4:      April 2013   Male Aged 16-17    E06000005     5
      5:        May 2013   Male Aged 16-17    E06000005     5
     ---
1034876:     August 2022 Female   Aged 65+ Column Total  9360
1034877: September 2022 Female   Aged 65+ Column Total  9225
1034878:    October 2022 Female   Aged 65+ Column Total  9225
1034879:   November 2022 Female   Aged 65+ Column Total  9315
1034880:   December 2022 Female   Aged 65+ Column Total  9495
```

```
>   dataset<-rbindlist(lapply(csv_files, function(x){fread(x,select= c("date","Gender","Age","geogcode","value"),sep = "\t",header=TRUE)}))
>   View(head(dataset,10),"Preview of unemployment dataset before transformation")
```

Fig 7.0 Preview of JSA dataset before transformation.

JSA PREVIEW BEFORE TRANSFORMATION

Sheffield Hallam University

EVIDENCE OF DATA QUALITY CHECK

```
>   print(data.table(miss_var_summary(dataset)))
   variable n_miss pct_miss
1:     date      0        0
2:      Age      0        0
3: geogcode      0        0
4:    value      0        0
>   # Create a plot of missing values before transformation
>   plot <- gg_miss_var(dataset,show_pct = TRUE)
>   print(plot + labs(title = "Plot of missing values in JSA dataset"))
```

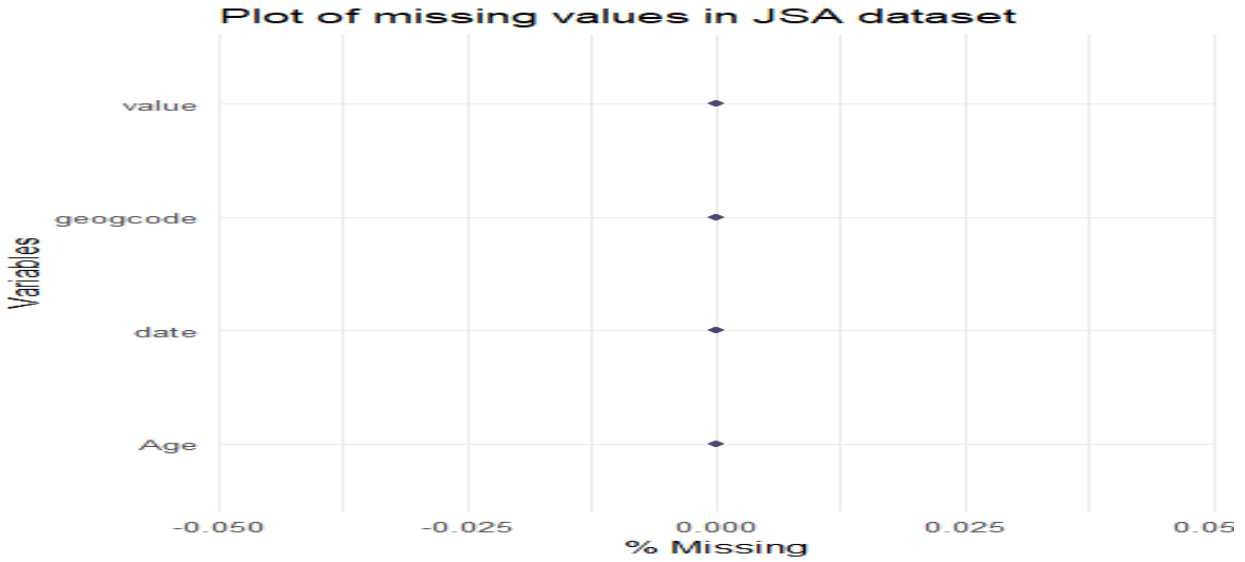


Fig 8.0: Evidence of missing values for Employment dataset

Job Seeker Allowance:

Invalid data values – There were no invalid data values in the dataset.

Missing values – There were no missing values.

Duplicate values – There were no duplicate values present.

```
# Data formatting: Extract and create month and year column from date column and make date column a key
#remove unnecessary strings from the Age column
dataset<-dataset[,.(year=format.Date(as.Date(paste0("01 ",date),format = "%d %B %Y"),"%Y"),
          month= format.Date(as.Date(paste0("01 ",date),format = "%d %B %Y"),"%m"),
          Gender,geogcode,Age=str_replace(Age, "Aged ", ""),value)]
dataset<-geography_code[,.GRP,.(County_code,County_name,Region_code,Region_name)][dataset,
                    on=.(County_code=geogcode),nomatch=0][,
                     .(value=sum(value)),keyby=.(year=as.numeric(year),month=as.numeric(month),Region_code,
```

Sheffield Hallam University

JSA PREVIEW AFTER TRANSFORMATION

	year	month	Region_code	Region_name	District_code	District_name	Age	Gender	value
1	2013	1	E12000001	North East	E06000001	Hartlepool	16-17	Female	0
2	2013	1	E12000001	North East	E06000001	Hartlepool	16-17	Male	0
3	2013	1	E12000001	North East	E06000001	Hartlepool	18-24	Female	465
4	2013	1	E12000001	North East	E06000001	Hartlepool	18-24	Male	960
5	2013	1	E12000001	North East	E06000001	Hartlepool	25-29	Female	220
6	2013	1	E12000001	North East	E06000001	Hartlepool	25-29	Male	550
7	2013	1	E12000001	North East	E06000001	Hartlepool	30-34	Female	155
8	2013	1	E12000001	North East	E06000001	Hartlepool	30-34	Male	375
9	2013	1	E12000001	North East	E06000001	Hartlepool	35-39	Female	145
10	2013	1	E12000001	North East	E06000001	Hartlepool	35-39	Male	305

Fig 9.0 Preview of JSA dataset after transformation.

Sheffield Hallam University

	Month	LSOA code	Crime type	Falls within
1	2010-12	E01017662	Other crime	Avon and Somerset Constabulary
2	2010-12	E01014399	Anti-social behaviour	Avon and Somerset Constabulary
3	2010-12	E01014399	Anti-social behaviour	Avon and Somerset Constabulary
4	2010-12	E01014399	Anti-social behaviour	Avon and Somerset Constabulary
5	2010-12	E01014399	Anti-social behaviour	Avon and Somerset Constabulary
6	2010-12	E01014399	Anti-social behaviour	Avon and Somerset Constabulary
7	2010-12	E01014399	Burglary	Avon and Somerset Constabulary
8	2010-12	E01014399	Burglary	Avon and Somerset Constabulary
9	2010-12	E01014399	Other crime	Avon and Somerset Constabulary
10	2010-12	E01014400	Anti-social behaviour	Avon and Somerset Constabulary

```
> crime_data
             Month LSOA code                 Crime type                 Falls within
       1: 2010-12 E01017662                Other crime Avon and Somerset Constabulary
       2: 2010-12 E01014399      Anti-social behaviour Avon and Somerset Constabulary
       3: 2010-12 E01014399      Anti-social behaviour Avon and Somerset Constabulary
       4: 2010-12 E01014399      Anti-social behaviour Avon and Somerset Constabulary
       5: 2010-12 E01014399      Anti-social behaviour Avon and Somerset Constabulary
      ---
75181930: 2022-12 E01031995              Public order                Wiltshire Police
75181931: 2022-12 E01031995                Shoplifting               Wiltshire Police
75181932: 2022-12 E01031995              Vehicle crime               Wiltshire Police
75181933: 2022-12 E01031995 Violence and sexual offences             Wiltshire Police
75181934: 2022-12 E01031995 Violence and sexual offences             Wiltshire Police
```

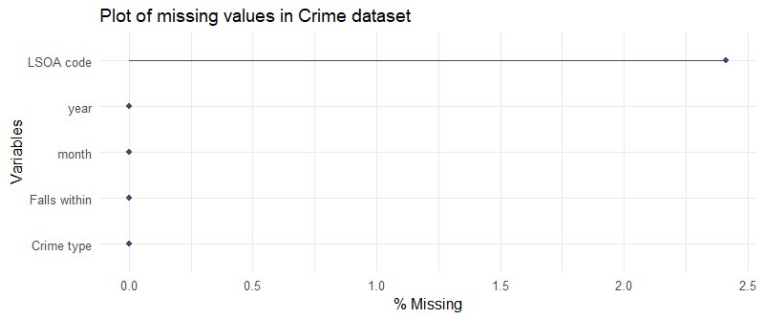
```
>   crime_data<-rbindlist(lapply(csv_files, function(x){fread(x,select= c("Month","LSOA code","Crime type","Falls within"))}))
> crime_data
```

Fig 10.0 Preview of Crime dataset before transformation.

CRIME DATASET PREVIEW BEFORE TRANSFORMATION

Sheffield Hallam University

EVIDENCE OF DATA QUALITY CHECK




```
print(data.table(miss_var_summary(crime_data)))
      variable  n_miss pct_miss
     LSOA code 1814374 2.413311
    Crime type       0 0.000000
   Falls within      0 0.000000
          year       0 0.000000
         month       0 0.000000
```

Fig 11.0: Evidence of missing values for crime dataset

Crime Dataset:

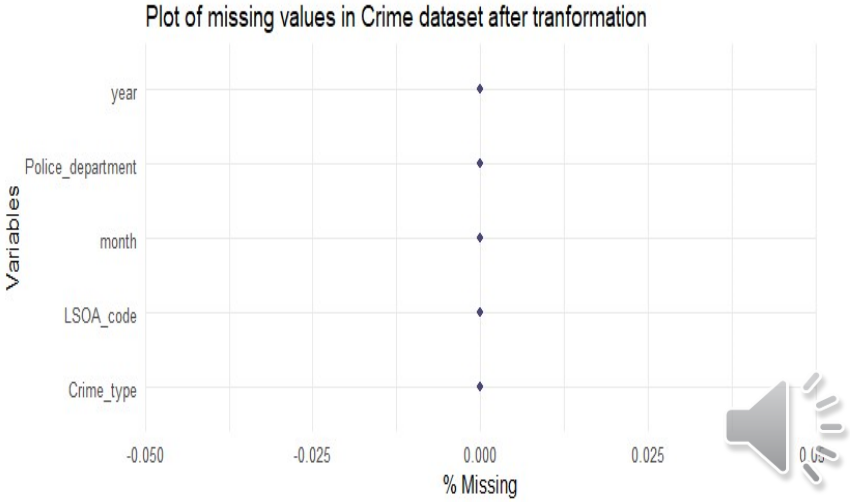
- Invalid data values – There were no invalid data values.
- Missing values – There were missing values, and the missing values were removed.
- Duplicate values – There were duplicate values in the dataset, and they were grouped together.
- Invalid date – No invalid date was found.

Exploratory data analysis using RStudio:

- Dropping of irrelevant columns
- Renaming columns
- Formatting Date

```
> #Data transformation
>
>   # Data formatting: Split the date column into month and year
>   crime_data[, c("year", "month") := tstrsplit(Month, "-", fixed=TRUE)][,Month:=NULL]
>
>   #filter out dates before 2013 as they are not needed in our business question
>   print(paste0(nrow(crime_data[year %in% c("2010","2011","2012")]),"count of datasets between year 2010 to 2012 before deleting"))
[1] "13328422count of datasets between year 2010 to 2012 before deleting"
>   crime_data<-crime_data[!(year %in% c("2010","2011","2012"))]
>
>   #rename the column names
>   setnames(crime_data,'LSOA code','LSOA_code')
>   setnames(crime_data,'Crime type','Crime_type')
>   setnames(crime_data,'Falls within','Police_department')
>
>   #Records with no LSOA code will be deleted
>   print(paste(nrow(crime_data[LSOA_code %in% c("",NA)]),"records have no LSOA code which will be deleted"))
[1] "2656791 records have no LSOA code which will be deleted"
>   crime_data<-crime_data[!(LSOA_code %in% c("",NA))]
```

Fig 12.0: Evidence showing missing values handled



Sheffield Hallam University

CRIME PREVIEW AFTER TRANSFORMATION

```
> crime_data
        year month Region_code Region_name        Police_department District_code         District_name                  Crime_type Total_crime
     1: 2013     1   E12000001  North East British Transport Police     E06000002         Middlesbrough         Anti-social behaviour           1
     2: 2013     1   E12000001  North East British Transport Police     E06000002         Middlesbrough                  Other theft           2
     3: 2013     1   E12000001  North East British Transport Police     E06000002         Middlesbrough   Public disorder and weapons           2
     4: 2013     1   E12000001  North East British Transport Police     E06000002         Middlesbrough                 Violent crime           1
     5: 2013     1   E12000001  North East British Transport Police     E06000003  Redcar and Cleveland                       Robbery           1
    ---
787400: 2022    12   W92000004       Wales      Warwickshire Police     W06000014 The Vale of Glamorgan                   Other crime           1
787401: 2022    12   W92000004       Wales       West Mercia Police     W06000020                Torfaen         Anti-social behaviour           1
787402: 2022    12   W92000004       Wales       West Mercia Police     W06000023                  Powys         Anti-social behaviour           1
787403: 2022    12   W92000004       Wales       West Mercia Police     W06000023                  Powys                       Burglary           1
787404: 2022    12   W92000004       Wales       West Mercia Police     W06000023                  Powys Violence and sexual offences           2
```

	year	month	Region_code	Region_name	Police_department	District_code	District_name	Crime_type	Total_crime
1	2013	1	E12000001	North East	British Transport Police	E06000002	Middlesbrough	Anti-social behaviour	1
2	2013	1	E12000001	North East	British Transport Police	E06000002	Middlesbrough	Other theft	2
3	2013	1	E12000001	North East	British Transport Police	E06000002	Middlesbrough	Public disorder and weapons	2
4	2013	1	E12000001	North East	British Transport Police	E06000002	Middlesbrough	Violent crime	1
5	2013	1	E12000001	North East	British Transport Police	E06000003	Redcar and Cleveland	Robbery	1
6	2013	1	E12000001	North East	British Transport Police	E06000004	Stockton-on-Tees	Anti-social behaviour	1
7	2013	1	E12000001	North East	British Transport Police	E06000004	Stockton-on-Tees	Public disorder and weapons	2
8	2013	1	E12000001	North East	British Transport Police	E06000005	Darlington	Anti-social behaviour	8
9	2013	1	E12000001	North East	British Transport Police	E06000005	Darlington	Criminal damage and arson	1
10	2013	1	E12000001	North East	British Transport Police	E06000005	Darlington	Other theft	6

Fig 13.0 Preview of Crime dataset after transformation.

Sheffield Hallam University

SCHEMA DIMENSION TABLES

```
>   CrimeTypeDim<<-crime_data[,.(CrimeTypeID=paste0('CR',.GRP+100)),Crime_type] #create custom crime_type ID as Dimension
>   print(CrimeTypeDim)
                    Crime_type CrimeTypeID
 1:        Anti-social behaviour       CR101
 2:                  Other theft       CR102
 3:  Public disorder and weapons       CR103
 4:                Violent crime       CR104
 5:                      Robbery       CR105
 6:    Criminal damage and arson       CR106
 7:                         Drugs       CR107
 8:                  Other crime       CR108
 9:                  Shoplifting       CR109
10:                     Burglary       CR110
11:                Vehicle crime       CR111
12:                 Public order       CR112
13:                Bicycle theft       CR113
14: Violence and sexual offences       CR114
15:        Possession of weapons       CR115
16:        Theft from the person       CR116
```

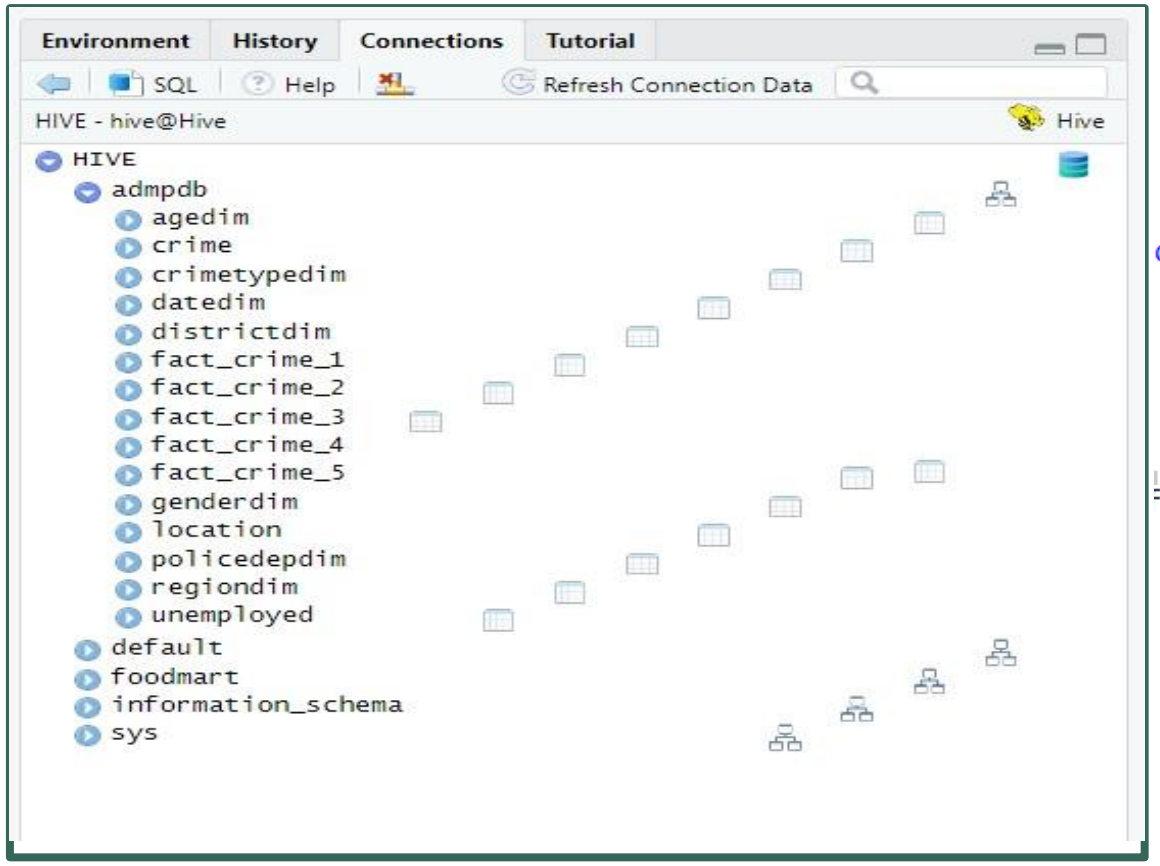
```
> DistrictDim<<-unique(geography_code[,.(District_code=County_code,District_name=County_name)]) #District Dimension
>  print(DistrictDim)
     District_code        District_name
  1:     E06000001           Hartlepool
  2:     E06000002          Middlesbrough
  3:     E06000003 Redcar and Cleveland
  4:     E06000004      Stockton-on-Tees
  5:     E06000005             Darlington
 ---
376:     W06000020               Torfaen
377:     W06000021          Monmouthshire
378:     W06000022                Newport
379:     W06000023                 Powys
380:     W06000024          Merthyr Tydfil
```

```
> DistrictDim<<-unique(geography_code[,.(District_code=County_code,District_name=County_name)]) #District Dimension
> print(dateDim)
     dateid year month
  1:  20131 2013     1
  2:  20132 2013     2
  3:  20133 2013     3
  4:  20134 2013     4
  5:  20135 2013     5
 ---
116:  20228 2022     8
117:  20229 2022     9
118: 202210 2022    10
119: 202211 2022    11
120: 202212 2022    12
```

Fig 14.0 Preview of Schema Dimension tables.

Sheffield Hallam University

EVIDENCE OF HIVE CONNECTION



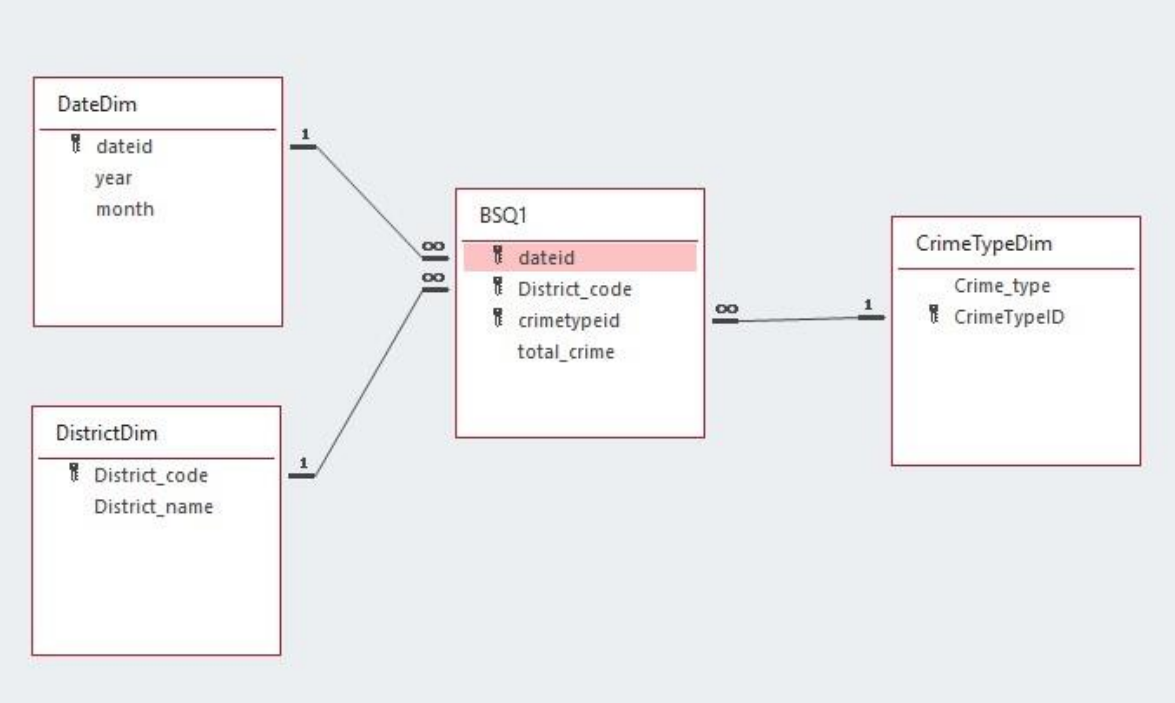

```
#Connect to Hive server
conhive <- dbConnect(odbc::odbc(),
                     Driver = "cloudera odbc driver for apache hive",
                     Host   = "sandbox-hdp.hortonworks.com",
                     UID    = "hive",
                     PWD    = "hive",
                     Port   = 10000)
data.table(dbGetQuery(conhive,"select * from admpdb.fact_crime_1 limit 5;"))
 fact_crime_1.year fact_crime_1.month fact_crime_1.district_code fact_crime_1
              2013                 1                 E06000001
              2013                 1                 E06000001
              2013                 1                 E06000001
              2013                 1                 E06000001
              2013                 1                 E06000001

act_crime_1.district_name   fact_crime_1.crime_type fact_crime_1.total_crime
               Hartlepool     Anti-social behaviour                      620
               Hartlepool                  Burglary                       80
               Hartlepool Criminal damage and arson                      123
               Hartlepool                     Drugs                       33
               Hartlepool              Other crime                       18
```

Fig 15.0 Evidence of hive connection.

Sheffield Hallam University

SCHEMA DIAGRAM



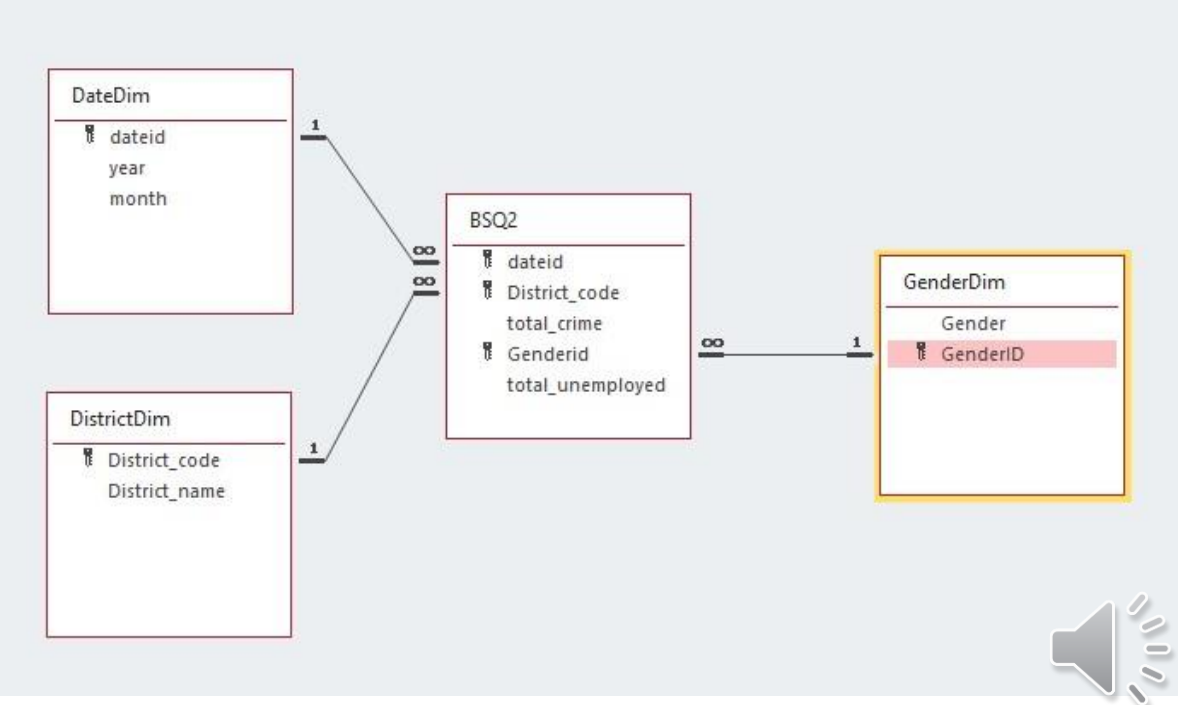
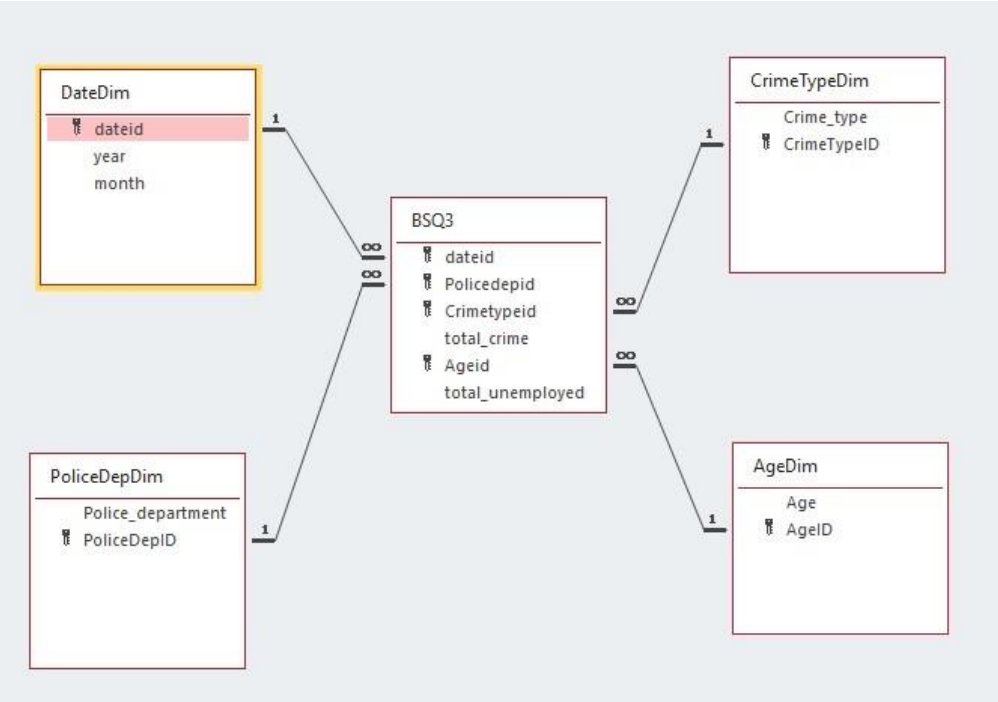


Fig 16.0, 17.0: Star schemas for business questions 1 and 2.

SCHEMA DIAGRAM CONTD



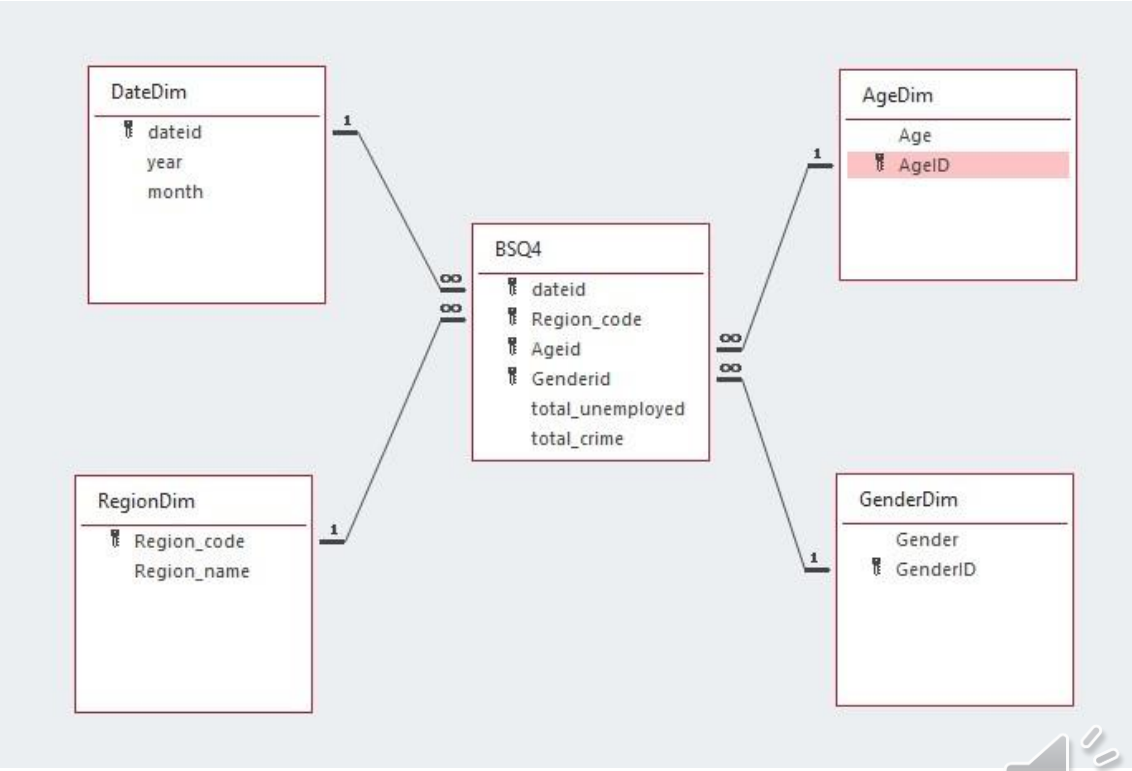
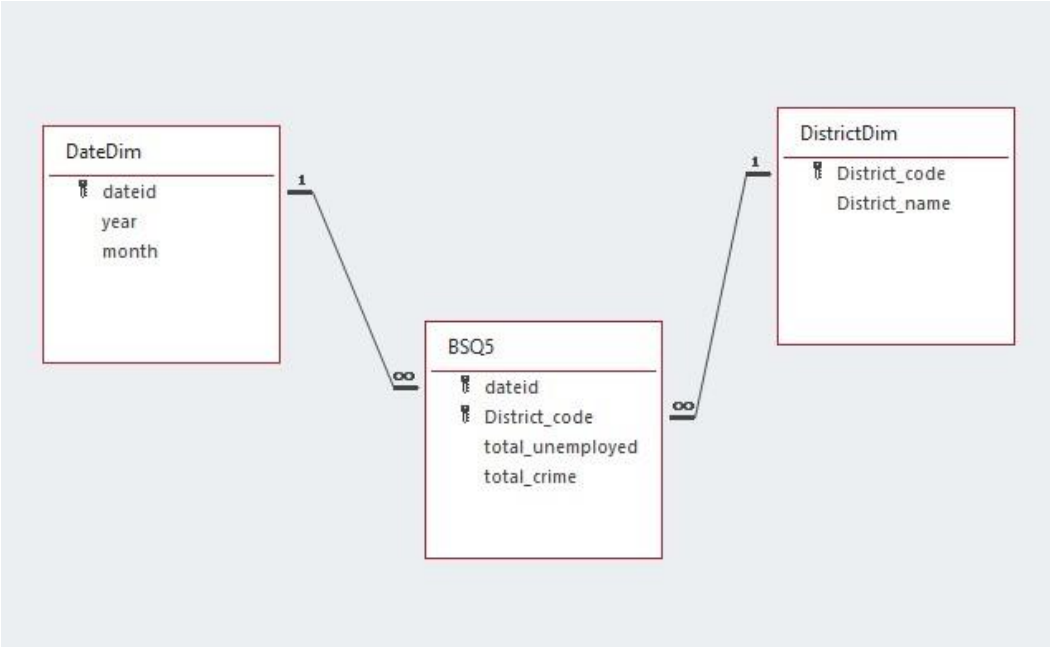


Fig 18.0 and 19.0: Star schemas for business questions 3 and 4.

SCHEMA DIAGRAM CONTD



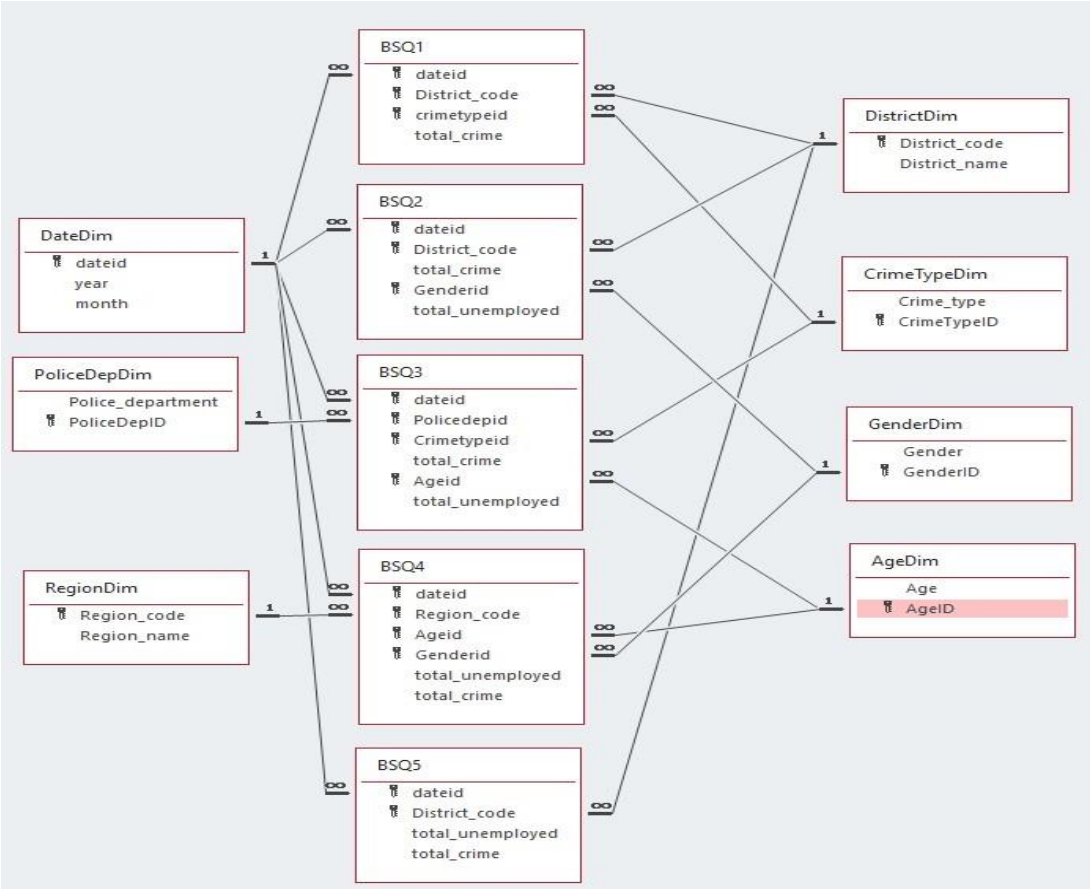


Fig 20.0 and 21.0: Star schemas for business questions 5 and Constellation

Sheffield Hallam University

AUTOMATED TABLEAU CONNECTION EVIDENCE

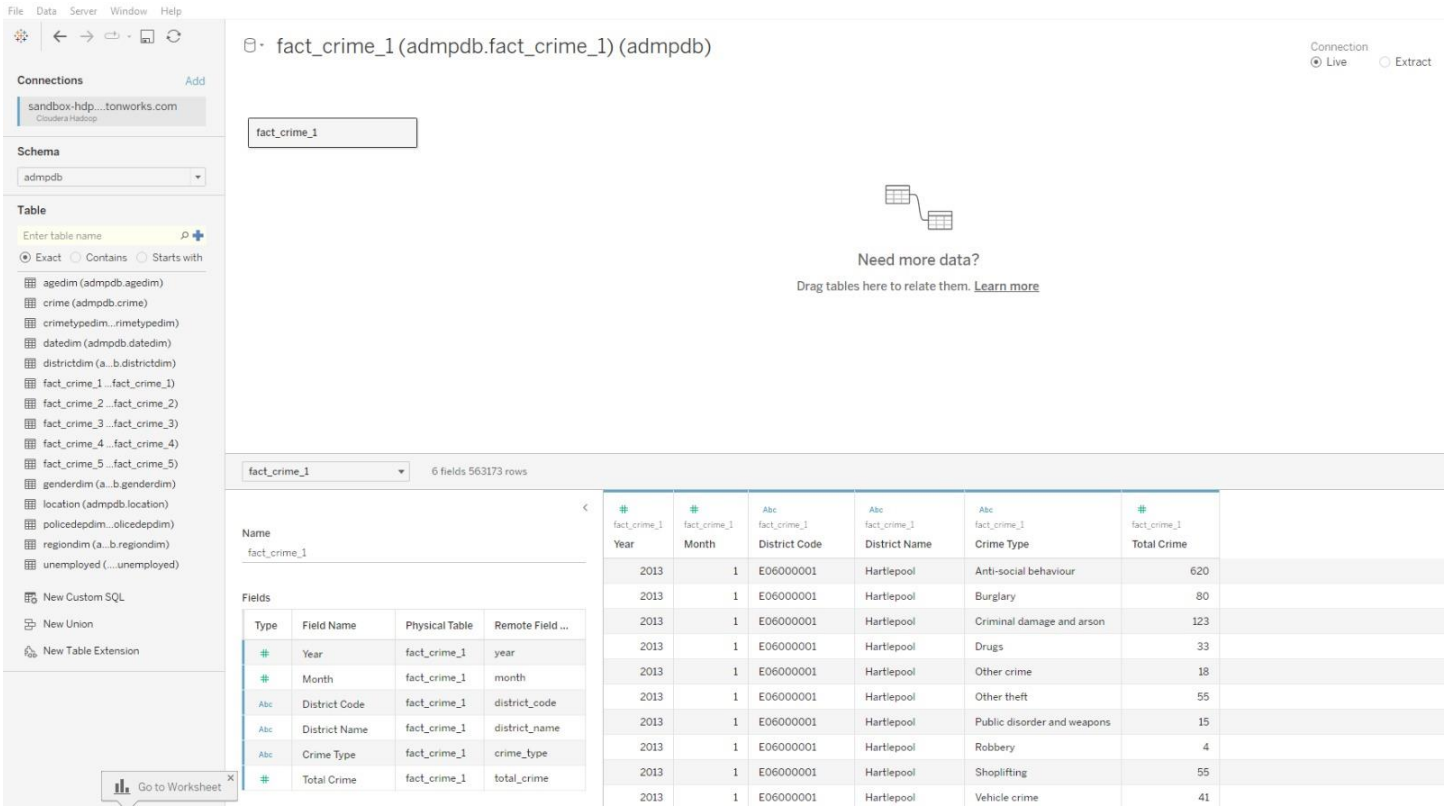


Fig 22.0 Evidence of tableau connection to Hive.

Sheffield Hallam University

ANSWER TO BUSINESS QUESTION 1

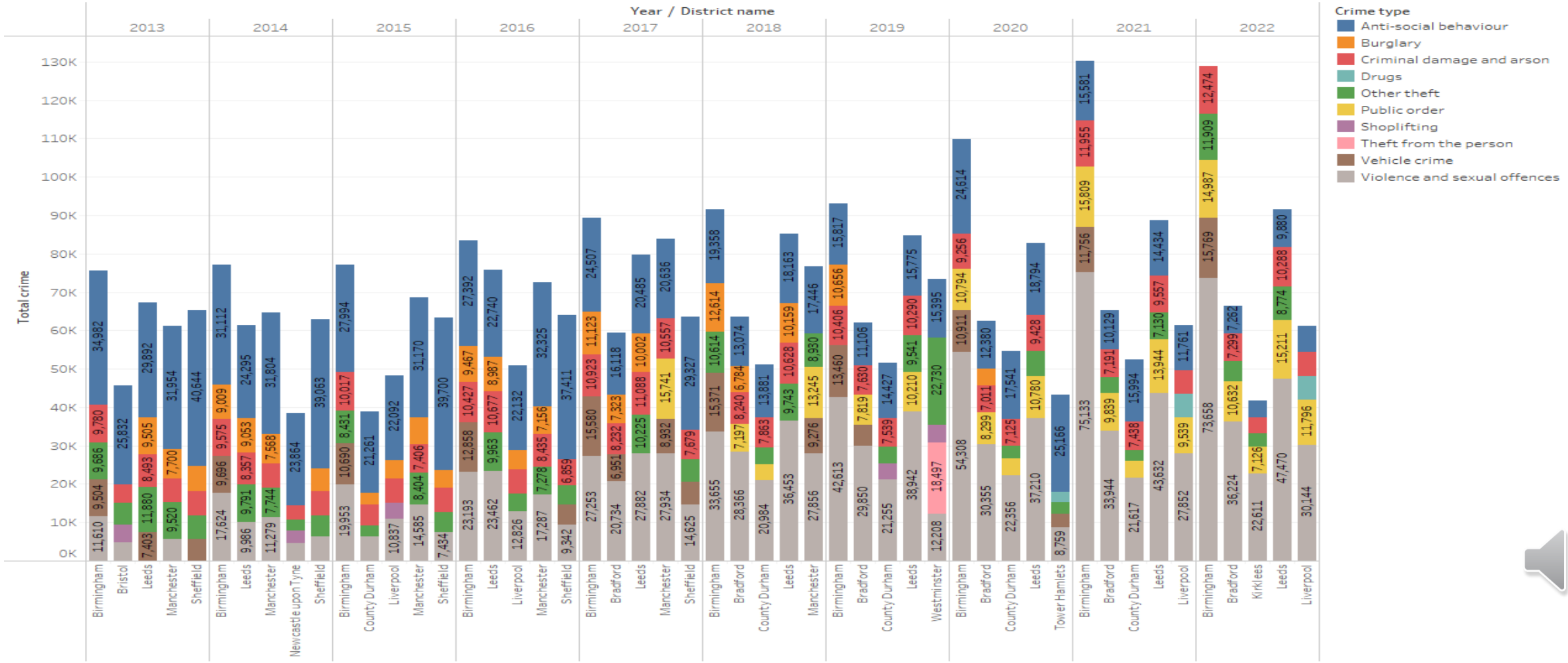


Fig 23.0: Pictorial representation of top 5 most crime prone location in UK per Crime type from 2013 to 2022. This reveals that Birmingham and Leeds are the most crime prone areas within the given period. The data also reveals that violence & sexual offences are the most frequent crime type.

Sheffield Hallam University

ANSWER TO BUSINESS QUESTION 1 CONTINUED

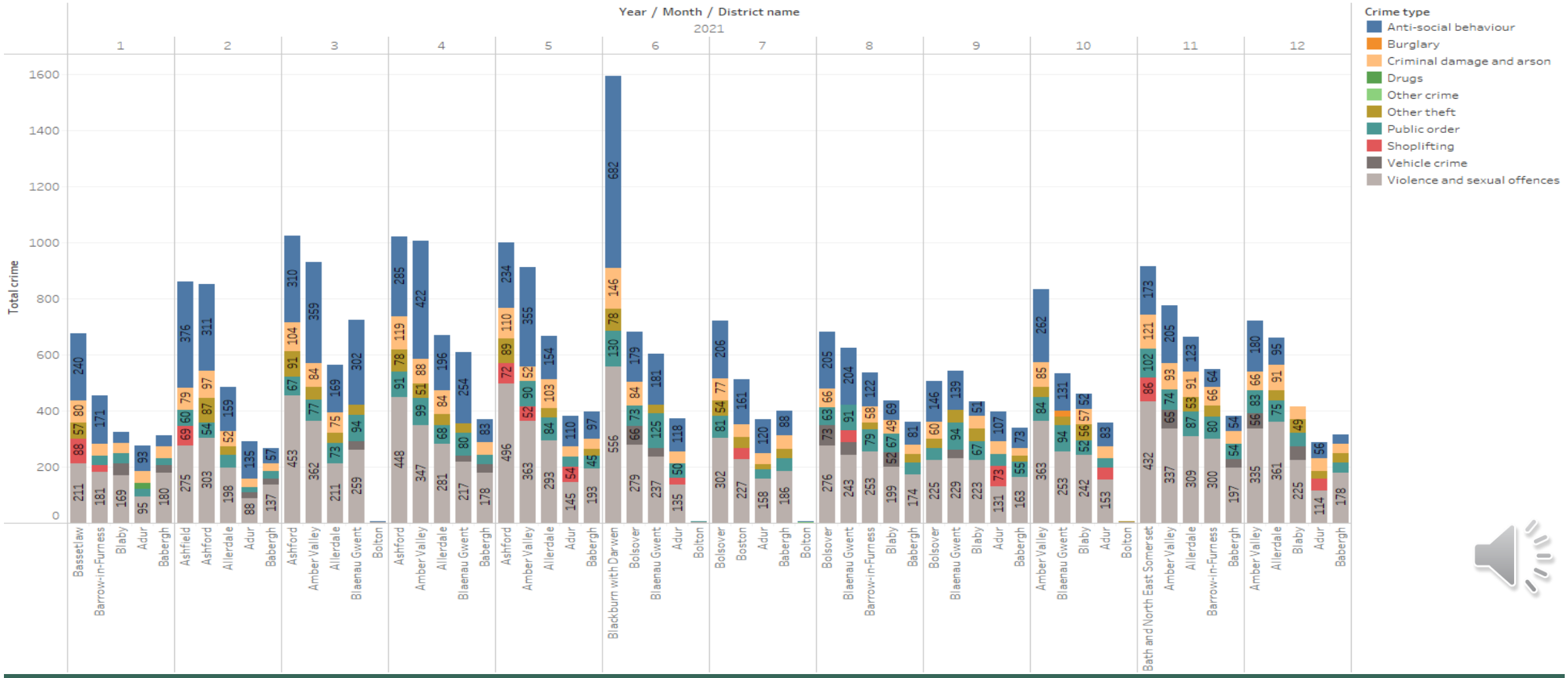


Fig 24.0: Pictorial representation of top 5 most crime prone location in UK per Crime type in the year 2021. This reveals that months between March to August records the highest number of crime with Anti-social behavior, Violence and Sexual crime taking the lead.

Sheffield Hallam University

ANSWER TO BUSINESS QUESTION 2

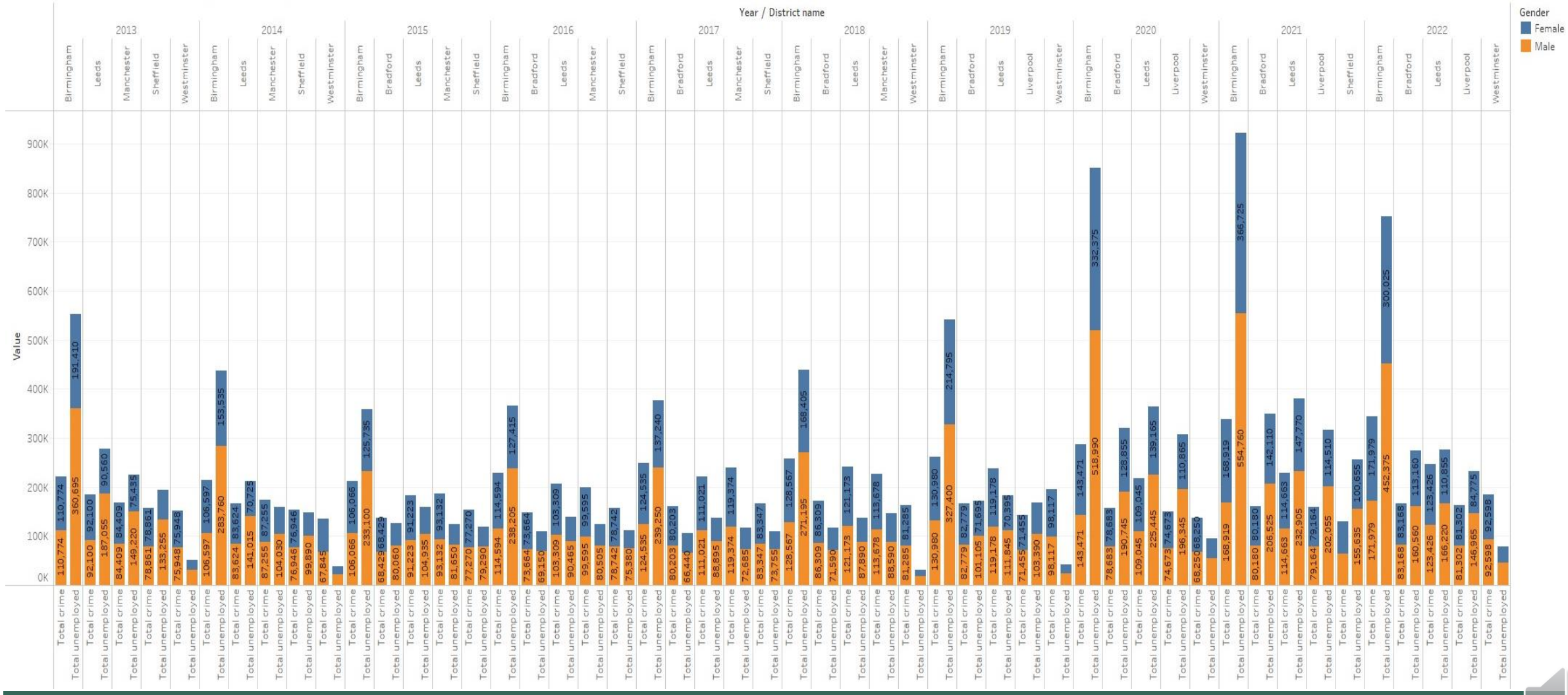


Fig 25.0: Pictorial representation of top 5 most crime prone location in UK per job seeker allowance gender between 2013 to 2022. This reveals that there is corresponding rise in crime whenever the male job seekers increases.

Advanced Data Management Project – Group 19

ANSWER TO BUSINESS QUESTION 3

Fig 26.0: Pictorial representation of top 5 Police force with the highest crime from 2013 to 2022. This reveals that Metropolitan Police force recorded the highest crime rate yearly while Antisocial behavior and Sexual offences remains the highest type of crime. The data also reveals that Age group 18-24 received the most Job seeker allowance.

Sheffield Hallam University

ANSWER TO BUSINESS QUESTION 4

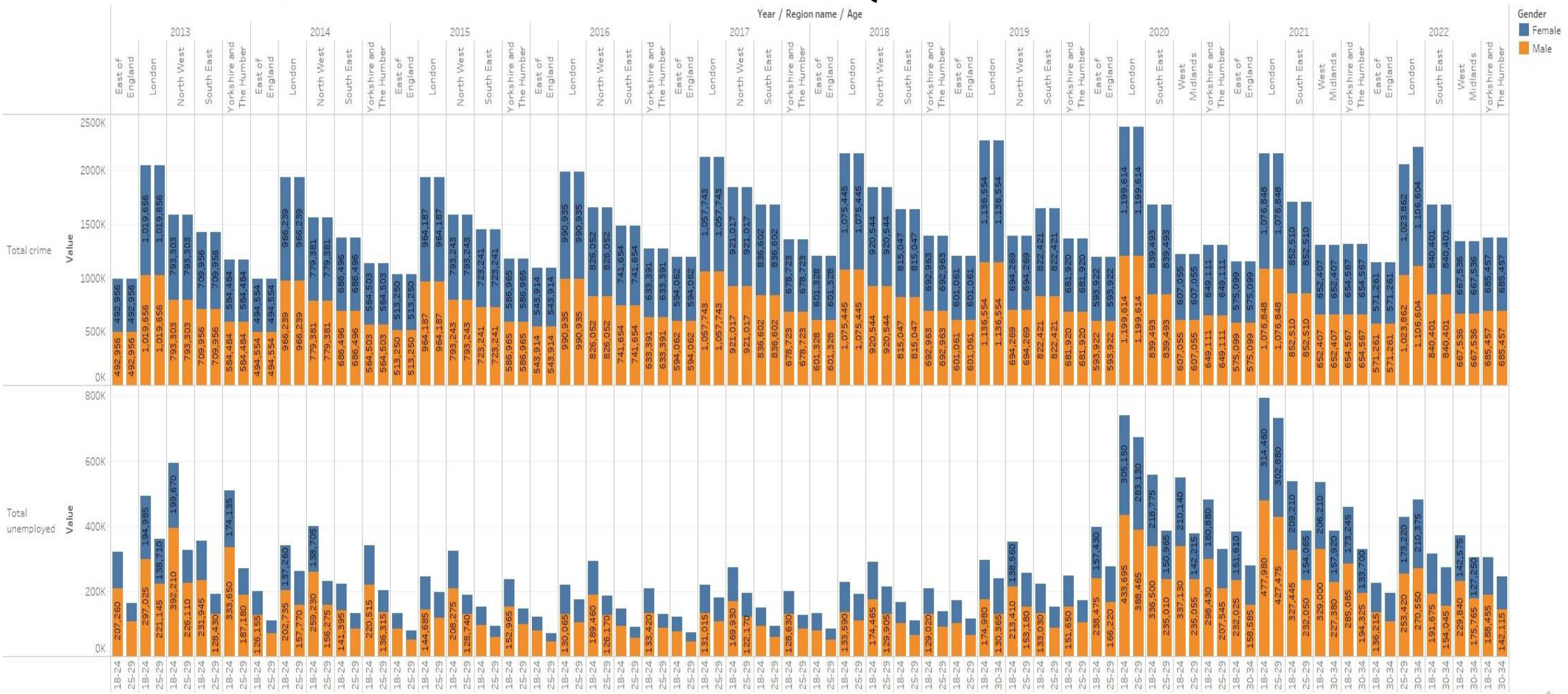


Fig 27.0: Pictorial representation of top 5 Regions in UK with the highest crime and the corresponding top two jobseeker allowance by age and gender from 2013 to 2022. The graph indicates that London Region has the highest crime rate and Age group 18-24 has the highest JSA.

Sheffield Hallam University

ANSWER TO BUSINESS QUESTION 5

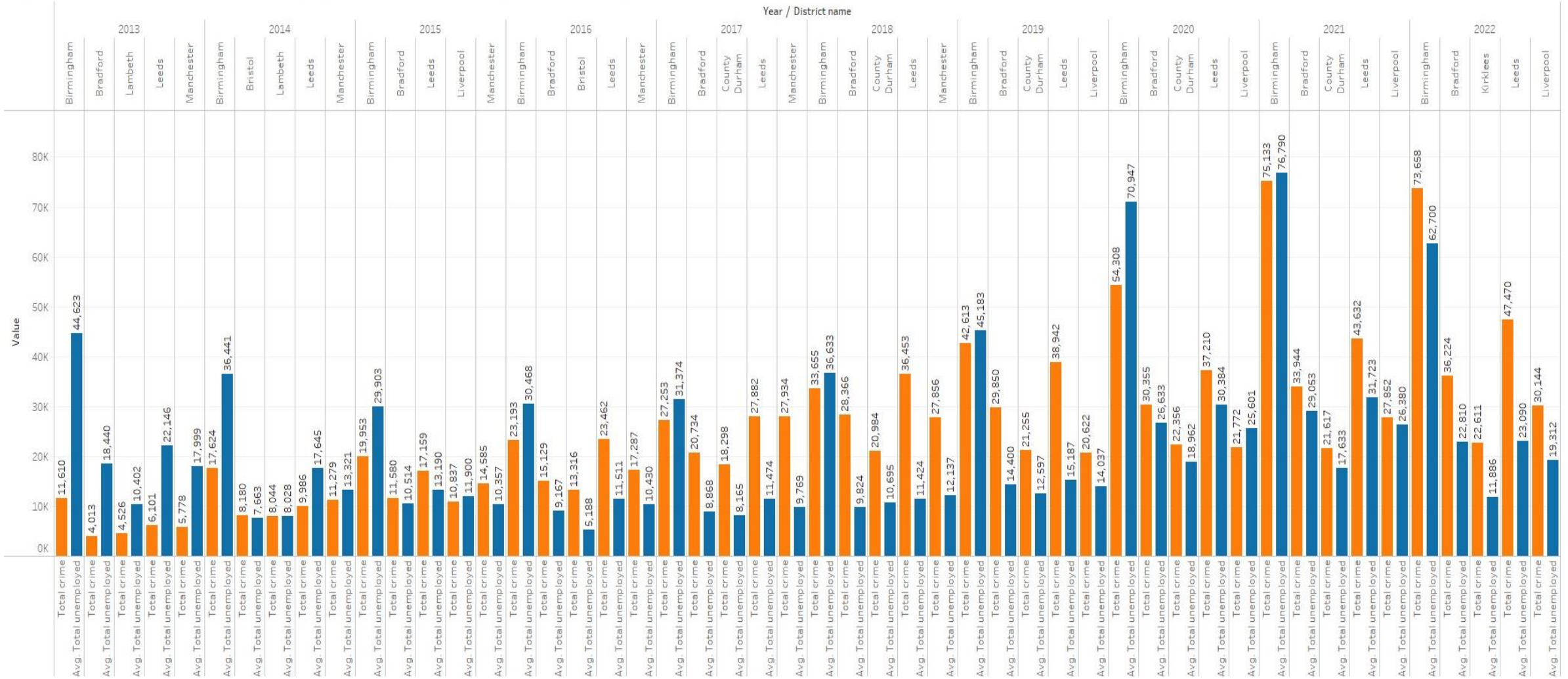


Fig 28.0: Pictorial representation of top 5 locations in UK with the highest crime caused by sexual violence and the corresponding jobseeker seeker from 2013 to 2022. The graph indicates that increase in Job seeker allowance give rise to increase in crime caused sexual violence.

Names	Roles	Responsibilities	Skills & Experience	Interdependency
Ebunoluwa Favour	Project Manager	Close monitoring of the project progress throughout its lifecycle and making necessary adjustments to ensure the project stays on track. Identifying and mitigating risks, resolving issues and managing changes to the project scope, schedule and budget.	She has 6 years working experience as a project manager with strong technical background in Big Data Technologies such as Hadoop, Apache Spark. Sound knowledge of project management methodologies like Agile and Waterfall.	SHU City council, Data Engineer, Business Analyst, Data Analyst, Data Quality Analyst, Data Visualization Specialist.
Saad Jeelani	Business Analyst	Understanding the business requirements and translating them into technical specifications that the team can implement. Provide valuable insights of the business domain, and market trends that will help the project team develop solution.	He has 5 years experience as a Business Analyst with solid understanding of Big Data tools including Hadoop, Spark and Hive. Strong problem-solving skills.	SHU Consult, Business Analyst, Data Analyst, Data Engineer, Data Quality Analyst, Data Visualization Specialist
Olatunji Ladokun	Data Engineer	Responsible for designing and implementing data architectures that are scalable, reliable and efficient. Development and maintaining data pipelines that extract, transform and load(ETL) data from various sources into a centralized data repository	Proficient in processing and storing large amount of data in distributed systems such as Hadoop, Spark and NoSQL. Strong programming skills in languages such as Python, R and SQL	SHU Consult, Business Analyst, Data Analyst, Data Engineer, Data Quality Analyst, Data Visualization Specialist
Mujeeb Ibrahim	Data Analyst	Responsible for data cleaning and preparation. Responsible for the data analysis, visualization, and communication of findings of the analysis effectively to stakeholders.	Proficient in statistical analysis and data mining techniques such as R, Access, Python. 4 years' experience in data analytic tools such as SQL, Tableau.	Data Visualization Specialist,Business Analyst, Data Analyst, Data Engineer, Data Quality Analyst, SHU Consult.
Kufre Mkpedem	Data Quality Analyst	Understanding the data, assessing its quality and identifying any anomalies, inconsistencies. Create and execute quality assurance tests to ensure the data meets the project's quality standards	Strong understanding of data quality policies and procedures. Proficient in data profiling techniques, statistical analysis and data governance.	SHU Consult, Business Analyst, Data Analyst, Data Engineer, Data Quality Analyst, Data Visualization Specialist
Joseph Anjaly	Data Visualization Analyst	Analyzing the data, identifying trends and patterns. Selecting the appropriate tools and software available based on complexity of the data, for creating the visualization	Possess strong understanding of data analytics tools such as Tableau. 5years experience of data modelling and warehousing concepts.	SHU Consult, Business Analyst, Data Analyst, Data Engineer, Data Quality Analyst, Data Visualization Specialist.

TEAM ROLES AND JUSTIFICATION

Sheffield Hallam University

PROJECT TIMELINE

SHU Consultancy

Mon, 03/07/2023

1

TASK	PROGRESS	START	END
Business Justification		3/7/23	6/7/23
Team Role Allocation		6/7/23	8/7/23
Requirement analysis		8/7/23	12/7/23
Data Sourcing and Pre-Processing			
Research and data sourcing		12/7/23	16/7/23
Data Collection/Extraction		14/7/23	19/7/23
Data preprocessing		19/7/23	22/7/23
Data Transformation		19/7/23	22/7/23
Data Warehousing			
Data Loading		22/7/23	27/7/23
Data Mart Creation		24/7/23	28/7/23
Data Validation and Quality Checks		28/7/23	2/8/23
Data Visualisation			
Tableau Cofiguration		2/8/23	3/8/23
Create Dashoards		3/8/23	10/8/23
Provide Recommedations		10/8/23	12/8/23
Project Completion			

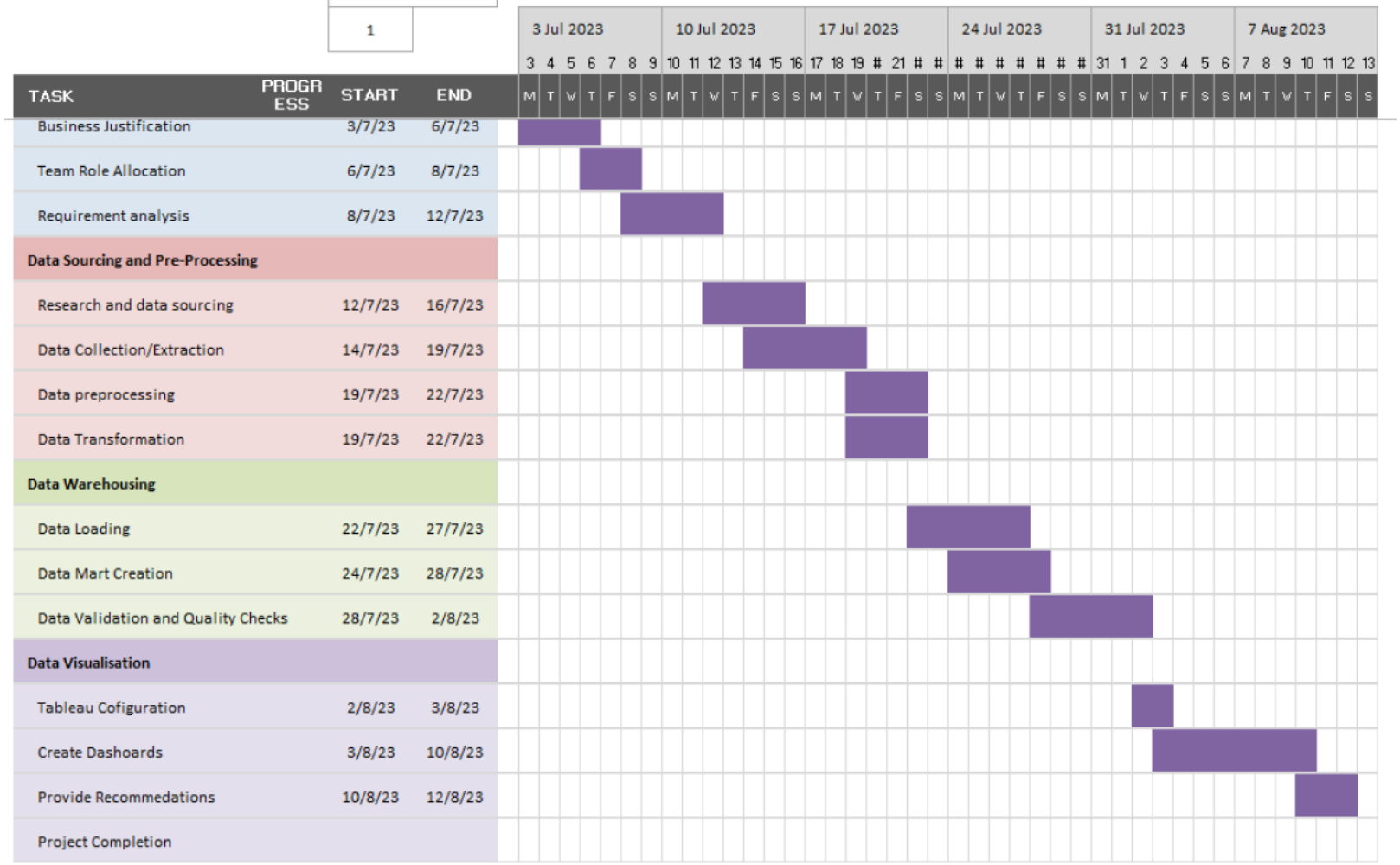


Fig 29.0 Time Plan

Sheffield Hallam University

CONCLUSSION

To sum up, the data showcased in the visuals highlights significant patterns in crime and societal indicators in the UK spanning the years 2013 to 2022. Birmingham and Leeds surface as areas with higher incidents of crime, predominantly marked by instances of violence and sexual offences. The period between March and August witnesses a noticeable uptick in criminal activities, notably involving anti-social conduct, violent incidents, and sexual offences. The observed connection between male job seekers and crime rates hints at a possible socioeconomic factor influencing criminal behaviour.

Consistently, the Metropolitan Police records the highest rate of criminal activity, underscoring the requirement for tailored law enforcement strategies. Additionally, the data underscores the importance of addressing allowances for job seekers, particularly within the 18-24 age bracket, as a potential measure to curb crime. Overall, an all-encompassing approach involving law enforcement, social welfare services, and community involvement emerges as imperative to combat crime and nurture safer neighbourhoods across the UK.

Sheffield Hallam University

RECOMMENDATION

Drawing from the analysis of the given data, it becomes apparent that specific regions within the UK have encountered elevated levels of criminal incidents, notably observed in Birmingham and Leeds.

The prominence of violent and sexual offenses further compounds these concerns. In response to these challenges, it is recommended that law enforcement agencies and local governing bodies deliberate on enacting focused crime deterrence approaches and community engagement initiatives tailored to these areas with heightened crime rates.

Also, fostering cooperation between the police, community groups, and social welfare services, it becomes feasible to delve into the underlying factors precipitating criminal activities and consequently cultivate more secure surroundings for the inhabitants.

Sheffield Hallam University

REFERENCES

- Altindag, D. T. (2012). Crime and unemployment: Evidence from Europe. *International Review of Law and Economics*, *32*(1), 145–157. https://doi.org/10.1016/j.irle.2011.10.003
- *Crime in United Kingdom and Wales: Police Force Area data tables - Office for National Statistics*. (2023, April 27). https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/policeforceareadatatables
- *Data downloads | data.police.uk*. (n.d.). https://data.police.uk/data/archive/
- *Dr Neil Chakraborti talks about hate crime*. (n.d.). [Video]. University of Leicester. https://le.ac.uk/hate-studies/research/the-leicester-hate-crime-project
- *Employment Rates by Ethnicity – London Datastore*. (n.d.). https://data.london.gov.uk/dataset/employment-rates-by-ethnicity
- Kinney, J. B., Brantingham, P. L., Wuschke, K., Kirk, M. S., & Brantingham, P. J. (2008). Crime Attractors, Generators and Detractors: Land Use and Urban Crime Opportunities. *Built Environment*, *34*(1), 62–74. https://doi.org/10.2148/benv.34.1.62
- *Recorded crime data at Police Force Area level (including pivot table) - Office for National Statistics*. (2019, April 25). https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/recordedcrimedataatpoliceforcearealevelincludingpivottable
- Tamplin, T. (2023). Bottom-Up Stock Analysis | Meaning, How It Works, Pros & Cons. *Finance Strategist*. https://www.financestrategists.com/wealth-management/fundamental-vs-technical-analysis/bottom-up-stock-analysis/

THANK YOU