Road Casualty Statistics Analysis

Your Name

2024-08-17

## Introduction

In this analysis, we explore the road casualty statistics for 2022. The dataset provides comprehensive data on various aspects of road accidents, including the number of vehicles involved, casualties, weather conditions, and more.

#### Step 1: Load and Inspect the Data

# Load necessary libraries  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readr)  
  
# Load the dataset  
data <- read\_csv("dft-road-casualty-statistics-collision-2022.csv")

## Warning: One or more parsing issues, call `problems()` on your data frame for details,  
## e.g.:  
## dat <- vroom(...)  
## problems(dat)

## Rows: 106004 Columns: 36  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (6): accident\_index, accident\_reference, date, local\_authority\_ons\_dis...  
## dbl (29): accident\_year, location\_easting\_osgr, location\_northing\_osgr, lon...  
## time (1): time  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Inspect the data  
glimpse(data)

## Rows: 106,004  
## Columns: 36  
## $ accident\_index <chr> "2022010352073", "20220103…  
## $ accident\_year <dbl> 2022, 2022, 2022, 2022, 20…  
## $ accident\_reference <chr> "010352073", "010352573", …  
## $ location\_easting\_osgr <dbl> 525199, 546214, 551119, 52…  
## $ location\_northing\_osgr <dbl> 177928, 179866, 174789, 19…  
## $ longitude <dbl> -0.198224, 0.105042, 0.173…  
## $ latitude <dbl> 51.48645, 51.49883, 51.451…  
## $ police\_force <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ accident\_severity <dbl> 3, 3, 3, 3, 3, 2, 3, 3, 3,…  
## $ number\_of\_vehicles <dbl> 2, 2, 2, 2, 4, 1, 2, 4, 1,…  
## $ number\_of\_casualties <dbl> 1, 1, 1, 2, 3, 5, 1, 4, 1,…  
## $ date <chr> "05/01/2022", "01/01/2022"…  
## $ day\_of\_week <dbl> 4, 7, 7, 7, 7, 7, 7, 7, 7,…  
## $ time <time> 16:40:00, 01:17:00, 01:15…  
## $ local\_authority\_district <dbl> -1, -1, -1, -1, -1, -1, -1…  
## $ local\_authority\_ons\_district <chr> "E09000013", "E09000011", …  
## $ local\_authority\_highway <chr> "E09000013", "E09000011", …  
## $ first\_road\_class <dbl> 3, 3, 3, 6, 3, 3, 6, 3, 3,…  
## $ first\_road\_number <dbl> 3218, 2016, 207, 0, 406, 1…  
## $ road\_type <dbl> 6, 3, 6, 6, 3, 6, 6, 3, 6,…  
## $ speed\_limit <dbl> 30, 50, 30, 30, 50, 30, 30…  
## $ junction\_detail <dbl> 6, 6, 9, 0, 5, 3, 3, 0, 7,…  
## $ junction\_control <dbl> 4, 2, 4, -1, 4, 4, 4, -1, …  
## $ second\_road\_class <dbl> 6, 6, 6, 0, 6, 6, 6, 0, 5,…  
## $ second\_road\_number <dbl> 0, 0, 0, -1, 0, 0, 0, -1, …  
## $ pedestrian\_crossing\_human\_control <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ pedestrian\_crossing\_physical\_facilities <dbl> 0, 4, 0, 0, 0, 0, 0, 0, 5,…  
## $ light\_conditions <dbl> 4, 4, 4, 4, 4, 4, 7, 4, 4,…  
## $ weather\_conditions <dbl> 1, 1, 1, 1, 1, 8, 1, 1, 1,…  
## $ road\_surface\_conditions <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ special\_conditions\_at\_site <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ carriageway\_hazards <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ urban\_or\_rural\_area <dbl> 1, 2, 1, 1, 1, 1, 1, 2, 1,…  
## $ did\_police\_officer\_attend\_scene\_of\_accident <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ trunk\_road\_flag <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2,…  
## $ lsoa\_of\_accident\_location <chr> "E01001883", "E01033745", …

#### Step 2: Identify and Drop Columns with Missing Values

# Identify columns with missing values  
missing\_data\_summary <- colSums(is.na(data))  
missing\_data\_summary

## accident\_index   
## 0   
## accident\_year   
## 0   
## accident\_reference   
## 0   
## location\_easting\_osgr   
## 22   
## location\_northing\_osgr   
## 22   
## longitude   
## 22   
## latitude   
## 22   
## police\_force   
## 0   
## accident\_severity   
## 0   
## number\_of\_vehicles   
## 0   
## number\_of\_casualties   
## 0   
## date   
## 0   
## day\_of\_week   
## 0   
## time   
## 0   
## local\_authority\_district   
## 0   
## local\_authority\_ons\_district   
## 0   
## local\_authority\_highway   
## 0   
## first\_road\_class   
## 0   
## first\_road\_number   
## 0   
## road\_type   
## 0   
## speed\_limit   
## 0   
## junction\_detail   
## 0   
## junction\_control   
## 0   
## second\_road\_class   
## 0   
## second\_road\_number   
## 0   
## pedestrian\_crossing\_human\_control   
## 0   
## pedestrian\_crossing\_physical\_facilities   
## 0   
## light\_conditions   
## 0   
## weather\_conditions   
## 0   
## road\_surface\_conditions   
## 0   
## special\_conditions\_at\_site   
## 0   
## carriageway\_hazards   
## 0   
## urban\_or\_rural\_area   
## 0   
## did\_police\_officer\_attend\_scene\_of\_accident   
## 0   
## trunk\_road\_flag   
## 0   
## lsoa\_of\_accident\_location   
## 0

# Drop columns with missing values  
data\_clean <- data %>%  
 select(-which(colSums(is.na(data)) > 0))  
  
# Inspect the cleaned dataset  
glimpse(data\_clean)

## Rows: 106,004  
## Columns: 32  
## $ accident\_index <chr> "2022010352073", "20220103…  
## $ accident\_year <dbl> 2022, 2022, 2022, 2022, 20…  
## $ accident\_reference <chr> "010352073", "010352573", …  
## $ police\_force <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ accident\_severity <dbl> 3, 3, 3, 3, 3, 2, 3, 3, 3,…  
## $ number\_of\_vehicles <dbl> 2, 2, 2, 2, 4, 1, 2, 4, 1,…  
## $ number\_of\_casualties <dbl> 1, 1, 1, 2, 3, 5, 1, 4, 1,…  
## $ date <chr> "05/01/2022", "01/01/2022"…  
## $ day\_of\_week <dbl> 4, 7, 7, 7, 7, 7, 7, 7, 7,…  
## $ time <time> 16:40:00, 01:17:00, 01:15…  
## $ local\_authority\_district <dbl> -1, -1, -1, -1, -1, -1, -1…  
## $ local\_authority\_ons\_district <chr> "E09000013", "E09000011", …  
## $ local\_authority\_highway <chr> "E09000013", "E09000011", …  
## $ first\_road\_class <dbl> 3, 3, 3, 6, 3, 3, 6, 3, 3,…  
## $ first\_road\_number <dbl> 3218, 2016, 207, 0, 406, 1…  
## $ road\_type <dbl> 6, 3, 6, 6, 3, 6, 6, 3, 6,…  
## $ speed\_limit <dbl> 30, 50, 30, 30, 50, 30, 30…  
## $ junction\_detail <dbl> 6, 6, 9, 0, 5, 3, 3, 0, 7,…  
## $ junction\_control <dbl> 4, 2, 4, -1, 4, 4, 4, -1, …  
## $ second\_road\_class <dbl> 6, 6, 6, 0, 6, 6, 6, 0, 5,…  
## $ second\_road\_number <dbl> 0, 0, 0, -1, 0, 0, 0, -1, …  
## $ pedestrian\_crossing\_human\_control <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ pedestrian\_crossing\_physical\_facilities <dbl> 0, 4, 0, 0, 0, 0, 0, 0, 5,…  
## $ light\_conditions <dbl> 4, 4, 4, 4, 4, 4, 7, 4, 4,…  
## $ weather\_conditions <dbl> 1, 1, 1, 1, 1, 8, 1, 1, 1,…  
## $ road\_surface\_conditions <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ special\_conditions\_at\_site <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ carriageway\_hazards <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ urban\_or\_rural\_area <dbl> 1, 2, 1, 1, 1, 1, 1, 2, 1,…  
## $ did\_police\_officer\_attend\_scene\_of\_accident <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ trunk\_road\_flag <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2,…  
## $ lsoa\_of\_accident\_location <chr> "E01001883", "E01033745", …

#### Step 3: Data Preparation

# Convert date to Date format and time to POSIXct  
data\_clean$date <- as.Date(data\_clean$date, format="%Y-%m-%d")  
data\_clean$time <- as.POSIXct(data\_clean$time, format="%H:%M")  
  
# Ensure all necessary columns are in the correct data format  
data\_clean <- data\_clean %>%  
 mutate(across(where(is.character), as.factor))

#### Step 4: Descriptive Statistics

# Summary statistics for key variables  
summary(data\_clean)

## accident\_index accident\_year accident\_reference police\_force   
## 2022010352073: 1 Min. :2022 010352073: 1 Min. : 1.0   
## 2022010352573: 1 1st Qu.:2022 010352573: 1 1st Qu.: 4.0   
## 2022010352575: 1 Median :2022 010352575: 1 Median :22.0   
## 2022010352578: 1 Mean :2022 010352578: 1 Mean :27.5   
## 2022010352580: 1 3rd Qu.:2022 010352580: 1 3rd Qu.:45.0   
## 2022010352588: 1 Max. :2022 010352588: 1 Max. :99.0   
## (Other) :105998 (Other) :105998   
## accident\_severity number\_of\_vehicles number\_of\_casualties date   
## Min. :1.00 Min. : 1.000 Min. : 1.000 Min. :NA   
## 1st Qu.:3.00 1st Qu.: 1.000 1st Qu.: 1.000 1st Qu.:NA   
## Median :3.00 Median : 2.000 Median : 1.000 Median :NA   
## Mean :2.75 Mean : 1.826 Mean : 1.278 Mean :NaN   
## 3rd Qu.:3.00 3rd Qu.: 2.000 3rd Qu.: 1.000 3rd Qu.:NA   
## Max. :3.00 Max. :16.000 Max. :16.000 Max. :NA   
## NA's :106004   
## day\_of\_week time local\_authority\_district  
## Min. :1.00 Min. :1970-01-01 00:00:00.0000 Min. :-1   
## 1st Qu.:3.00 1st Qu.:1970-01-01 10:30:00.0000 1st Qu.:-1   
## Median :4.00 Median :1970-01-01 14:55:00.0000 Median :-1   
## Mean :4.17 Mean :1970-01-01 14:09:52.6525 Mean :-1   
## 3rd Qu.:6.00 3rd Qu.:1970-01-01 17:59:00.0000 3rd Qu.:-1   
## Max. :7.00 Max. :1970-01-01 23:59:00.0000 Max. :-1   
##   
## local\_authority\_ons\_district local\_authority\_highway first\_road\_class  
## E08000025: 2383 E10000016: 3457 Min. :1.000   
## E08000035: 1585 E10000030: 2783 1st Qu.:3.000   
## E09000033: 1274 E08000025: 2383 Median :4.000   
## E09000022: 1167 E10000012: 2377 Mean :4.222   
## E08000032: 1062 E10000014: 2278 3rd Qu.:6.000   
## E09000030: 1039 E10000017: 2159 Max. :6.000   
## (Other) :97494 (Other) :90567   
## first\_road\_number road\_type speed\_limit junction\_detail   
## Min. : 0.0 Min. :1.000 Min. :20.00 Min. : 0.000   
## 1st Qu.: 0.0 1st Qu.:6.000 1st Qu.:30.00 1st Qu.: 0.000   
## Median : 31.0 Median :6.000 Median :30.00 Median : 2.000   
## Mean : 784.8 Mean :5.253 Mean :35.97 Mean : 4.016   
## 3rd Qu.: 533.0 3rd Qu.:6.000 3rd Qu.:40.00 3rd Qu.: 3.000   
## Max. :9176.0 Max. :9.000 Max. :70.00 Max. :99.000   
##   
## junction\_control second\_road\_class second\_road\_number  
## Min. :-1.000 Min. :-1.000 Min. : -1.0   
## 1st Qu.:-1.000 1st Qu.: 0.000 1st Qu.: -1.0   
## Median : 2.000 Median : 3.000 Median : 0.0   
## Mean : 1.715 Mean : 3.064 Mean : 217.9   
## 3rd Qu.: 4.000 3rd Qu.: 6.000 3rd Qu.: 0.0   
## Max. : 9.000 Max. : 6.000 Max. :9170.0   
##   
## pedestrian\_crossing\_human\_control pedestrian\_crossing\_physical\_facilities  
## Min. :-1.0000 Min. :-1.00   
## 1st Qu.: 0.0000 1st Qu.: 0.00   
## Median : 0.0000 Median : 0.00   
## Mean : 0.3464 Mean : 1.15   
## 3rd Qu.: 0.0000 3rd Qu.: 0.00   
## Max. : 9.0000 Max. : 9.00   
##   
## light\_conditions weather\_conditions road\_surface\_conditions  
## Min. :-1.000 Min. :1.000 Min. :-1.000   
## 1st Qu.: 1.000 1st Qu.:1.000 1st Qu.: 1.000   
## Median : 1.000 Median :1.000 Median : 1.000   
## Mean : 2.011 Mean :1.637 Mean : 1.346   
## 3rd Qu.: 4.000 3rd Qu.:1.000 3rd Qu.: 2.000   
## Max. : 7.000 Max. :9.000 Max. : 9.000   
##   
## special\_conditions\_at\_site carriageway\_hazards urban\_or\_rural\_area  
## Min. :-1.0000 Min. :-1.0000 Min. :1.000   
## 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.:1.000   
## Median : 0.0000 Median : 0.0000 Median :1.000   
## Mean : 0.2422 Mean : 0.1975 Mean :1.323   
## 3rd Qu.: 0.0000 3rd Qu.: 0.0000 3rd Qu.:2.000   
## Max. : 9.0000 Max. : 9.0000 Max. :3.000   
##   
## did\_police\_officer\_attend\_scene\_of\_accident trunk\_road\_flag   
## Min. :1.000 Min. :-1.000   
## 1st Qu.:1.000 1st Qu.: 2.000   
## Median :1.000 Median : 2.000   
## Mean :1.482 Mean : 1.726   
## 3rd Qu.:2.000 3rd Qu.: 2.000   
## Max. :3.000 Max. : 2.000   
##   
## lsoa\_of\_accident\_location  
## -1 : 4148   
## E01004736: 123   
## E01032739: 110   
## E01033708: 63   
## E01033595: 59   
## E01030458: 55   
## (Other) :101446

# Frequency of accidents by severity  
accident\_severity\_dist <- data\_clean %>%  
 group\_by(accident\_severity) %>%  
 summarise(count = n())  
  
accident\_severity\_dist

## # A tibble: 3 × 2  
## accident\_severity count  
## <dbl> <int>  
## 1 1 1602  
## 2 2 23333  
## 3 3 81069

#### Step 5: Analyze Factors Contributing to Accident Severity

# Accident severity by urban or rural area  
severity\_by\_area <- data\_clean %>%  
 group\_by(urban\_or\_rural\_area, accident\_severity) %>%  
 summarise(count = n())

## `summarise()` has grouped output by 'urban\_or\_rural\_area'. You can override  
## using the `.groups` argument.

severity\_by\_area

## # A tibble: 8 × 3  
## # Groups: urban\_or\_rural\_area [3]  
## urban\_or\_rural\_area accident\_severity count  
## <dbl> <dbl> <int>  
## 1 1 1 580  
## 2 1 2 14431  
## 3 1 3 56752  
## 4 2 1 1022  
## 5 2 2 8896  
## 6 2 3 24301  
## 7 3 2 6  
## 8 3 3 16

# Impact of weather conditions on the number of casualties  
weather\_vs\_casualties <- data\_clean %>%  
 group\_by(weather\_conditions) %>%  
 summarise(avg\_casualties = mean(number\_of\_casualties))  
  
weather\_vs\_casualties

## # A tibble: 9 × 2  
## weather\_conditions avg\_casualties  
## <dbl> <dbl>  
## 1 1 1.28  
## 2 2 1.30  
## 3 3 1.34  
## 4 4 1.34  
## 5 5 1.35  
## 6 6 1.43  
## 7 7 1.28  
## 8 8 1.24  
## 9 9 1.11

#### Step 6: Analyze Impact of Time on Accident Frequency and Severity

# Create a new column for the hour of the day  
data\_clean <- data\_clean %>%  
 mutate(hour = format(time, "%H"))  
  
# Calculate average severity per hour  
time\_vs\_severity <- data\_clean %>%  
 group\_by(hour) %>%  
 summarise(avg\_severity = mean(accident\_severity))  
  
time\_vs\_severity

## # A tibble: 24 × 2  
## hour avg\_severity  
## <chr> <dbl>  
## 1 00 2.64  
## 2 01 2.66  
## 3 02 2.66  
## 4 03 2.65  
## 5 04 2.66  
## 6 05 2.69  
## 7 06 2.72  
## 8 07 2.75  
## 9 08 2.80  
## 10 09 2.77  
## # ℹ 14 more rows

#### Step 7: Logistic Regression for Predicting Accident Severity

# Convert accident severity to binary (1 for fatal, 0 for non-fatal)  
data\_clean <- data\_clean %>%  
 mutate(severe\_accident = ifelse(accident\_severity == 1, 1, 0))  
  
# Logistic regression model  
logistic\_model <- glm(severe\_accident ~ weather\_conditions + road\_surface\_conditions + hour,   
 data = data\_clean, family = binomial)  
  
summary(logistic\_model)

##   
## Call:  
## glm(formula = severe\_accident ~ weather\_conditions + road\_surface\_conditions +   
## hour, family = binomial, data = data\_clean)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.09874 0.13033 -23.777 < 2e-16 \*\*\*  
## weather\_conditions -0.09442 0.01852 -5.099 3.42e-07 \*\*\*  
## road\_surface\_conditions 0.03621 0.02985 1.213 0.225171   
## hour01 0.01475 0.19128 0.077 0.938529   
## hour02 0.06437 0.20866 0.308 0.757712   
## hour03 0.22872 0.20737 1.103 0.270034   
## hour04 0.10919 0.22403 0.487 0.625983   
## hour05 0.04404 0.20685 0.213 0.831405   
## hour06 -0.56206 0.19645 -2.861 0.004223 \*\*   
## hour07 -1.30625 0.19314 -6.763 1.35e-11 \*\*\*  
## hour08 -1.63713 0.18343 -8.925 < 2e-16 \*\*\*  
## hour09 -1.20253 0.18276 -6.580 4.71e-11 \*\*\*  
## hour10 -0.95606 0.17046 -5.609 2.04e-08 \*\*\*  
## hour11 -1.07096 0.16988 -6.304 2.90e-10 \*\*\*  
## hour12 -1.17868 0.16880 -6.983 2.89e-12 \*\*\*  
## hour13 -1.16054 0.16687 -6.955 3.54e-12 \*\*\*  
## hour14 -1.08220 0.16199 -6.681 2.38e-11 \*\*\*  
## hour15 -1.25295 0.15988 -7.837 4.62e-15 \*\*\*  
## hour16 -1.19848 0.15803 -7.584 3.36e-14 \*\*\*  
## hour17 -1.40772 0.16256 -8.660 < 2e-16 \*\*\*  
## hour18 -1.35564 0.16867 -8.037 9.17e-16 \*\*\*  
## hour19 -0.99441 0.16388 -6.068 1.30e-09 \*\*\*  
## hour20 -0.87526 0.17161 -5.100 3.39e-07 \*\*\*  
## hour21 -0.66350 0.17355 -3.823 0.000132 \*\*\*  
## hour22 -0.49216 0.17026 -2.891 0.003845 \*\*   
## hour23 -0.39632 0.17644 -2.246 0.024695 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 16612 on 106003 degrees of freedom  
## Residual deviance: 16247 on 105978 degrees of freedom  
## AIC: 16299  
##   
## Number of Fisher Scoring iterations: 7

#### Step 8: Prepare Data for Power BI

# Export the cleaned dataset for further analysis in Power BI  
write\_csv(data\_clean, "cleaned\_road\_accident\_data.csv")  
  
# Export the summarized data for visualization  
write\_csv(accident\_severity\_dist, "accident\_severity\_dist.csv")  
write\_csv(weather\_vs\_casualties, "weather\_vs\_casualties.csv")  
write\_csv(time\_vs\_severity, "time\_vs\_severity.csv")

##### Interaction Effects in Logistic Regression

# Interaction effect in logistic regression  
interaction\_model <- glm(severe\_accident ~ weather\_conditions \* road\_surface\_conditions + hour,   
 data = data\_clean, family = binomial)  
  
summary(interaction\_model)

##   
## Call:  
## glm(formula = severe\_accident ~ weather\_conditions \* road\_surface\_conditions +   
## hour, family = binomial, data = data\_clean)  
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -3.210062 0.137652 -23.320  
## weather\_conditions -0.055477 0.023988 -2.313  
## road\_surface\_conditions 0.111729 0.041176 2.713  
## hour01 0.013360 0.191301 0.070  
## hour02 0.065189 0.208681 0.312  
## hour03 0.225380 0.207382 1.087  
## hour04 0.108529 0.224057 0.484  
## hour05 0.037536 0.206902 0.181  
## hour06 -0.571646 0.196511 -2.909  
## hour07 -1.309933 0.193162 -6.782  
## hour08 -1.637398 0.183439 -8.926  
## hour09 -1.202510 0.182768 -6.579  
## hour10 -0.953591 0.170474 -5.594  
## hour11 -1.066873 0.169897 -6.280  
## hour12 -1.172750 0.168822 -6.947  
## hour13 -1.154021 0.166908 -6.914  
## hour14 -1.075493 0.162022 -6.638  
## hour15 -1.246745 0.159912 -7.796  
## hour16 -1.193931 0.158055 -7.554  
## hour17 -1.405214 0.162568 -8.644  
## hour18 -1.353471 0.168677 -8.024  
## hour19 -0.992536 0.163895 -6.056  
## hour20 -0.873473 0.171619 -5.090  
## hour21 -0.663332 0.173563 -3.822  
## hour22 -0.494222 0.170277 -2.902  
## hour23 -0.395010 0.176457 -2.239  
## weather\_conditions:road\_surface\_conditions -0.020530 0.008966 -2.290  
## Pr(>|z|)   
## (Intercept) < 2e-16 \*\*\*  
## weather\_conditions 0.020739 \*   
## road\_surface\_conditions 0.006659 \*\*   
## hour01 0.944322   
## hour02 0.754747   
## hour03 0.277132   
## hour04 0.628117   
## hour05 0.856039   
## hour06 0.003626 \*\*   
## hour07 1.19e-11 \*\*\*  
## hour08 < 2e-16 \*\*\*  
## hour09 4.72e-11 \*\*\*  
## hour10 2.22e-08 \*\*\*  
## hour11 3.40e-10 \*\*\*  
## hour12 3.74e-12 \*\*\*  
## hour13 4.71e-12 \*\*\*  
## hour14 3.18e-11 \*\*\*  
## hour15 6.37e-15 \*\*\*  
## hour16 4.22e-14 \*\*\*  
## hour17 < 2e-16 \*\*\*  
## hour18 1.02e-15 \*\*\*  
## hour19 1.40e-09 \*\*\*  
## hour20 3.59e-07 \*\*\*  
## hour21 0.000132 \*\*\*  
## hour22 0.003702 \*\*   
## hour23 0.025184 \*   
## weather\_conditions:road\_surface\_conditions 0.022041 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 16612 on 106003 degrees of freedom  
## Residual deviance: 16241 on 105977 degrees of freedom  
## AIC: 16295  
##   
## Number of Fisher Scoring iterations: 7

##### Feature Engineering

# Create Day/Night indicator  
data\_clean <- data\_clean %>%  
 mutate(day\_night = ifelse(hour >= 6 & hour < 18, "Day", "Night"))  
  
# Create Weekend indicator  
data\_clean <- data\_clean %>%  
 mutate(weekend = ifelse(day\_of\_week %in% c(6, 7), "Weekend", "Weekday"))

##### Clustering Analysis

# Select relevant features for clustering  
cluster\_data <- data\_clean %>%  
 select(number\_of\_casualties, number\_of\_vehicles, weather\_conditions, road\_surface\_conditions) %>%  
 scale()  
  
# Perform k-means clustering  
set.seed(123)  
kmeans\_result <- kmeans(cluster\_data, centers = 3, nstart = 20)  
  
# Add cluster information to the dataset  
data\_clean$cluster <- kmeans\_result$cluster

##### Analyzing Temporal Trends

# Create a month and season column  
data\_clean <- data\_clean %>%  
 mutate(month = format(date, "%m"),  
 season = case\_when(  
 month %in% c("12", "01", "02") ~ "Winter",  
 month %in% c("03", "04", "05") ~ "Spring",  
 month %in% c("06", "07", "08") ~ "Summer",  
 TRUE ~ "Fall"  
 ))  
  
# Analyze accident frequency by season  
accidents\_by\_season <- data\_clean %>%  
 group\_by(season) %>%  
 summarise(count = n(), avg\_severity = mean(accident\_severity))  
  
accidents\_by\_season

## # A tibble: 1 × 3  
## season count avg\_severity  
## <chr> <int> <dbl>  
## 1 Fall 106004 2.75

##### Correlation

# Install the corrplot package if not already installed  
if (!require("corrplot")) {  
 install.packages("corrplot", dependencies = TRUE)  
 library(corrplot)  
} else {  
 library(corrplot)  
}

## Loading required package: corrplot

## corrplot 0.94 loaded

# Check if any plotting device is open and close it  
if (dev.cur() != 1) dev.off()

## null device   
## 1

# Filter out non-numeric columns  
data\_numeric <- data\_clean %>%  
 select\_if(is.numeric)  
  
# Check if all columns are numeric  
str(data\_numeric)

## tibble [106,004 × 27] (S3: tbl\_df/tbl/data.frame)  
## $ accident\_year : num [1:106004] 2022 2022 2022 2022 2022 ...  
## $ police\_force : num [1:106004] 1 1 1 1 1 1 1 1 1 1 ...  
## $ accident\_severity : num [1:106004] 3 3 3 3 3 2 3 3 3 2 ...  
## $ number\_of\_vehicles : num [1:106004] 2 2 2 2 4 1 2 4 1 1 ...  
## $ number\_of\_casualties : num [1:106004] 1 1 1 2 3 5 1 4 1 1 ...  
## $ day\_of\_week : num [1:106004] 4 7 7 7 7 7 7 7 7 7 ...  
## $ local\_authority\_district : num [1:106004] -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...  
## $ first\_road\_class : num [1:106004] 3 3 3 6 3 3 6 3 3 4 ...  
## $ first\_road\_number : num [1:106004] 3218 2016 207 0 406 ...  
## $ road\_type : num [1:106004] 6 3 6 6 3 6 6 3 6 6 ...  
## $ speed\_limit : num [1:106004] 30 50 30 30 50 30 30 40 30 20 ...  
## $ junction\_detail : num [1:106004] 6 6 9 0 5 3 3 0 7 3 ...  
## $ junction\_control : num [1:106004] 4 2 4 -1 4 4 4 -1 2 4 ...  
## $ second\_road\_class : num [1:106004] 6 6 6 0 6 6 6 0 5 6 ...  
## $ second\_road\_number : num [1:106004] 0 0 0 -1 0 0 0 -1 0 0 ...  
## $ pedestrian\_crossing\_human\_control : num [1:106004] 0 0 0 0 0 0 0 0 0 0 ...  
## $ pedestrian\_crossing\_physical\_facilities : num [1:106004] 0 4 0 0 0 0 0 0 5 1 ...  
## $ light\_conditions : num [1:106004] 4 4 4 4 4 4 7 4 4 5 ...  
## $ weather\_conditions : num [1:106004] 1 1 1 1 1 8 1 1 1 1 ...  
## $ road\_surface\_conditions : num [1:106004] 1 1 1 1 1 1 1 1 1 1 ...  
## $ special\_conditions\_at\_site : num [1:106004] 0 0 0 0 0 0 0 0 0 0 ...  
## $ carriageway\_hazards : num [1:106004] 0 0 0 0 0 0 0 0 0 0 ...  
## $ urban\_or\_rural\_area : num [1:106004] 1 2 1 1 1 1 1 2 1 1 ...  
## $ did\_police\_officer\_attend\_scene\_of\_accident: num [1:106004] 1 1 1 1 1 1 1 1 1 1 ...  
## $ trunk\_road\_flag : num [1:106004] 2 2 2 2 2 2 2 2 2 2 ...  
## $ severe\_accident : num [1:106004] 0 0 0 0 0 0 0 0 0 0 ...  
## $ cluster : int [1:106004] 1 1 1 1 2 2 1 2 1 1 ...

# Calculate the correlation matrix  
correlation\_matrix <- cor(data\_numeric, use = "complete.obs")

## Warning in cor(data\_numeric, use = "complete.obs"): the standard deviation is  
## zero

# Display the correlation matrix  
print(correlation\_matrix)

## accident\_year police\_force  
## accident\_year 1 NA  
## police\_force NA 1.000000000  
## accident\_severity NA -0.080583592  
## number\_of\_vehicles NA -0.006955904  
## number\_of\_casualties NA 0.062257633  
## day\_of\_week NA 0.001057650  
## local\_authority\_district NA NA  
## first\_road\_class NA 0.028425340  
## first\_road\_number NA 0.010109997  
## road\_type NA 0.002737158  
## speed\_limit NA 0.291471846  
## junction\_detail NA -0.173773305  
## junction\_control NA -0.162420060  
## second\_road\_class NA -0.148203067  
## second\_road\_number NA -0.021142131  
## pedestrian\_crossing\_human\_control NA -0.213333554  
## pedestrian\_crossing\_physical\_facilities NA -0.214809833  
## light\_conditions NA 0.002539140  
## weather\_conditions NA -0.071024180  
## road\_surface\_conditions NA -0.050040596  
## special\_conditions\_at\_site NA -0.123744405  
## carriageway\_hazards NA -0.121134195  
## urban\_or\_rural\_area NA 0.295660891  
## did\_police\_officer\_attend\_scene\_of\_accident NA -0.182403388  
## trunk\_road\_flag NA -0.621024523  
## severe\_accident NA 0.045141865  
## cluster NA -0.042038631  
## accident\_severity  
## accident\_year NA  
## police\_force -0.0805835919  
## accident\_severity 1.0000000000  
## number\_of\_vehicles 0.0632737197  
## number\_of\_casualties -0.0966158140  
## day\_of\_week -0.0008345923  
## local\_authority\_district NA  
## first\_road\_class 0.0153739762  
## first\_road\_number -0.0224625526  
## road\_type -0.0288261217  
## speed\_limit -0.1049018868  
## junction\_detail 0.0655780208  
## junction\_control 0.0772211710  
## second\_road\_class 0.0618547302  
## second\_road\_number 0.0201327522  
## pedestrian\_crossing\_human\_control 0.0726511371  
## pedestrian\_crossing\_physical\_facilities 0.0532159788  
## light\_conditions -0.0520089420  
## weather\_conditions 0.0417939499  
## road\_surface\_conditions 0.0188910768  
## special\_conditions\_at\_site 0.0418042941  
## carriageway\_hazards 0.0363851647  
## urban\_or\_rural\_area -0.1025088509  
## did\_police\_officer\_attend\_scene\_of\_accident 0.1904222144  
## trunk\_road\_flag 0.0730154170  
## severe\_accident -0.4643089019  
## cluster 0.0032937135  
## number\_of\_vehicles  
## accident\_year NA  
## police\_force -0.006955904  
## accident\_severity 0.063273720  
## number\_of\_vehicles 1.000000000  
## number\_of\_casualties 0.212343278  
## day\_of\_week -0.002242139  
## local\_authority\_district NA  
## first\_road\_class -0.124929853  
## first\_road\_number -0.002266323  
## road\_type -0.066448482  
## speed\_limit 0.090614027  
## junction\_detail 0.003821251  
## junction\_control 0.037735958  
## second\_road\_class 0.037665602  
## second\_road\_number 0.008088767  
## pedestrian\_crossing\_human\_control -0.003417427  
## pedestrian\_crossing\_physical\_facilities -0.045457006  
## light\_conditions -0.065735078  
## weather\_conditions -0.039831850  
## road\_surface\_conditions -0.030019039  
## special\_conditions\_at\_site -0.005585140  
## carriageway\_hazards -0.021493947  
## urban\_or\_rural\_area 0.042968365  
## did\_police\_officer\_attend\_scene\_of\_accident -0.010457309  
## trunk\_road\_flag -0.008037724  
## severe\_accident -0.008872215  
## cluster 0.134210549  
## number\_of\_casualties day\_of\_week  
## accident\_year NA NA  
## police\_force 0.0622576326 0.0010576501  
## accident\_severity -0.0966158140 -0.0008345923  
## number\_of\_vehicles 0.2123432785 -0.0022421386  
## number\_of\_casualties 1.0000000000 0.0001969107  
## day\_of\_week 0.0001969107 1.0000000000  
## local\_authority\_district NA NA  
## first\_road\_class -0.0815016340 0.0054311264  
## first\_road\_number 0.0148898082 -0.0022591402  
## road\_type -0.0238238603 0.0047584391  
## speed\_limit 0.1835550791 -0.0105459203  
## junction\_detail -0.0442843793 0.0044057930  
## junction\_control -0.0545679542 0.0045736252  
## second\_road\_class -0.0487325553 0.0048445070  
## second\_road\_number 0.0016826762 0.0008405925  
## pedestrian\_crossing\_human\_control -0.0542831758 0.0016488820  
## pedestrian\_crossing\_physical\_facilities -0.0636188616 -0.0011572460  
## light\_conditions 0.0375239899 0.0049912472  
## weather\_conditions -0.0303567260 -0.0057873177  
## road\_surface\_conditions -0.0085895625 -0.0046302108  
## special\_conditions\_at\_site -0.0329782376 -0.0032169815  
## carriageway\_hazards -0.0362420673 -0.0024202902  
## urban\_or\_rural\_area 0.1440040327 -0.0055150870  
## did\_police\_officer\_attend\_scene\_of\_accident -0.1261971171 -0.0011499794  
## trunk\_road\_flag -0.0581187467 0.0006439579  
## severe\_accident 0.0672717209 0.0002434010  
## cluster 0.3073152453 -0.0014406954  
## local\_authority\_district  
## accident\_year NA  
## police\_force NA  
## accident\_severity NA  
## number\_of\_vehicles NA  
## number\_of\_casualties NA  
## day\_of\_week NA  
## local\_authority\_district 1  
## first\_road\_class NA  
## first\_road\_number NA  
## road\_type NA  
## speed\_limit NA  
## junction\_detail NA  
## junction\_control NA  
## second\_road\_class NA  
## second\_road\_number NA  
## pedestrian\_crossing\_human\_control NA  
## pedestrian\_crossing\_physical\_facilities NA  
## light\_conditions NA  
## weather\_conditions NA  
## road\_surface\_conditions NA  
## special\_conditions\_at\_site NA  
## carriageway\_hazards NA  
## urban\_or\_rural\_area NA  
## did\_police\_officer\_attend\_scene\_of\_accident NA  
## trunk\_road\_flag NA  
## severe\_accident NA  
## cluster NA  
## first\_road\_class first\_road\_number  
## accident\_year NA NA  
## police\_force 0.028425340 0.010109997  
## accident\_severity 0.015373976 -0.022462553  
## number\_of\_vehicles -0.124929853 -0.002266323  
## number\_of\_casualties -0.081501634 0.014889808  
## day\_of\_week 0.005431126 -0.002259140  
## local\_authority\_district NA NA  
## first\_road\_class 1.000000000 -0.258102653  
## first\_road\_number -0.258102653 1.000000000  
## road\_type 0.286207352 -0.004170758  
## speed\_limit -0.323831702 -0.015917035  
## junction\_detail 0.019679297 -0.025167192  
## junction\_control 0.073991905 0.009767843  
## second\_road\_class 0.090647060 0.028313558  
## second\_road\_number -0.051740993 0.128260375  
## pedestrian\_crossing\_human\_control -0.037476913 -0.032201641  
## pedestrian\_crossing\_physical\_facilities -0.108747343 0.027107397  
## light\_conditions -0.037443153 0.004184592  
## weather\_conditions 0.008133094 -0.002755031  
## road\_surface\_conditions -0.014065825 -0.008112434  
## special\_conditions\_at\_site -0.020955881 -0.023547258  
## carriageway\_hazards -0.016006460 -0.024693916  
## urban\_or\_rural\_area -0.179821372 0.002215652  
## did\_police\_officer\_attend\_scene\_of\_accident 0.045718076 -0.038372190  
## trunk\_road\_flag 0.132077072 -0.015972574  
## severe\_accident -0.033717958 0.007862445  
## cluster -0.032184856 -0.002904445  
## road\_type speed\_limit  
## accident\_year NA NA  
## police\_force 0.0027371583 0.291471846  
## accident\_severity -0.0288261217 -0.104901887  
## number\_of\_vehicles -0.0664484824 0.090614027  
## number\_of\_casualties -0.0238238603 0.183555079  
## day\_of\_week 0.0047584391 -0.010545920  
## local\_authority\_district NA NA  
## first\_road\_class 0.2862073519 -0.323831702  
## first\_road\_number -0.0041707584 -0.015917035  
## road\_type 1.0000000000 -0.179285802  
## speed\_limit -0.1792858024 1.000000000  
## junction\_detail 0.1152797381 -0.150654956  
## junction\_control -0.0034714388 -0.281126865  
## second\_road\_class 0.0288752129 -0.312949925  
## second\_road\_number -0.0951468808 -0.027346813  
## pedestrian\_crossing\_human\_control 0.1162913096 -0.153019840  
## pedestrian\_crossing\_physical\_facilities 0.0228363983 -0.250591065  
## light\_conditions -0.0063354788 0.084679281  
## weather\_conditions 0.0519257206 -0.041759182  
## road\_surface\_conditions 0.0802982960 0.006315599  
## special\_conditions\_at\_site 0.1170243431 -0.066288267  
## carriageway\_hazards 0.1225681186 -0.066847695  
## urban\_or\_rural\_area -0.0487343793 0.673260664  
## did\_police\_officer\_attend\_scene\_of\_accident 0.0554486421 -0.226355179  
## trunk\_road\_flag 0.0623080734 -0.246165859  
## severe\_accident 0.0006919176 0.096138353  
## cluster 0.0344227741 0.036808507  
## junction\_detail junction\_control  
## accident\_year NA NA  
## police\_force -0.173773305 -0.162420060  
## accident\_severity 0.065578021 0.077221171  
## number\_of\_vehicles 0.003821251 0.037735958  
## number\_of\_casualties -0.044284379 -0.054567954  
## day\_of\_week 0.004405793 0.004573625  
## local\_authority\_district NA NA  
## first\_road\_class 0.019679297 0.073991905  
## first\_road\_number -0.025167192 0.009767843  
## road\_type 0.115279738 -0.003471439  
## speed\_limit -0.150654956 -0.281126865  
## junction\_detail 1.000000000 0.448784543  
## junction\_control 0.448784543 1.000000000  
## second\_road\_class 0.263824065 0.888098826  
## second\_road\_number 0.010380315 0.142460985  
## pedestrian\_crossing\_human\_control 0.362599666 0.168109305  
## pedestrian\_crossing\_physical\_facilities 0.263016395 0.172148522  
## light\_conditions -0.003406128 -0.056285314  
## weather\_conditions 0.102914596 0.041029294  
## road\_surface\_conditions 0.153693228 0.064112913  
## special\_conditions\_at\_site 0.287368813 0.112068771  
## carriageway\_hazards 0.280231862 0.107270812  
## urban\_or\_rural\_area -0.131103860 -0.226458176  
## did\_police\_officer\_attend\_scene\_of\_accident 0.257163580 0.127507315  
## trunk\_road\_flag 0.061110207 0.108237735  
## severe\_accident -0.024431454 -0.055599957  
## cluster 0.080967892 0.008917945  
## second\_road\_class  
## accident\_year NA  
## police\_force -0.148203067  
## accident\_severity 0.061854730  
## number\_of\_vehicles 0.037665602  
## number\_of\_casualties -0.048732555  
## day\_of\_week 0.004844507  
## local\_authority\_district NA  
## first\_road\_class 0.090647060  
## first\_road\_number 0.028313558  
## road\_type 0.028875213  
## speed\_limit -0.312949925  
## junction\_detail 0.263824065  
## junction\_control 0.888098826  
## second\_road\_class 1.000000000  
## second\_road\_number 0.036972215  
## pedestrian\_crossing\_human\_control 0.089959499  
## pedestrian\_crossing\_physical\_facilities 0.160896869  
## light\_conditions -0.057815201  
## weather\_conditions 0.017208955  
## road\_surface\_conditions 0.009848056  
## special\_conditions\_at\_site 0.041169295  
## carriageway\_hazards 0.033321629  
## urban\_or\_rural\_area -0.264616413  
## did\_police\_officer\_attend\_scene\_of\_accident 0.076890111  
## trunk\_road\_flag 0.118124686  
## severe\_accident -0.053093796  
## cluster -0.012586111  
## second\_road\_number  
## accident\_year NA  
## police\_force -0.0211421308  
## accident\_severity 0.0201327522  
## number\_of\_vehicles 0.0080887671  
## number\_of\_casualties 0.0016826762  
## day\_of\_week 0.0008405925  
## local\_authority\_district NA  
## first\_road\_class -0.0517409931  
## first\_road\_number 0.1282603753  
## road\_type -0.0951468808  
## speed\_limit -0.0273468132  
## junction\_detail 0.0103803149  
## junction\_control 0.1424609851  
## second\_road\_class 0.0369722153  
## second\_road\_number 1.0000000000  
## pedestrian\_crossing\_human\_control -0.0048586453  
## pedestrian\_crossing\_physical\_facilities 0.0723323732  
## light\_conditions 0.0005423747  
## weather\_conditions -0.0015321515  
## road\_surface\_conditions -0.0072207121  
## special\_conditions\_at\_site -0.0107103488  
## carriageway\_hazards -0.0119307819  
## urban\_or\_rural\_area -0.0223959461  
## did\_police\_officer\_attend\_scene\_of\_accident 0.0022344871  
## trunk\_road\_flag 0.0014039466  
## severe\_accident -0.0148863514  
## cluster -0.0061147663  
## pedestrian\_crossing\_human\_control  
## accident\_year NA  
## police\_force -0.213333554  
## accident\_severity 0.072651137  
## number\_of\_vehicles -0.003417427  
## number\_of\_casualties -0.054283176  
## day\_of\_week 0.001648882  
## local\_authority\_district NA  
## first\_road\_class -0.037476913  
## first\_road\_number -0.032201641  
## road\_type 0.116291310  
## speed\_limit -0.153019840  
## junction\_detail 0.362599666  
## junction\_control 0.168109305  
## second\_road\_class 0.089959499  
## second\_road\_number -0.004858645  
## pedestrian\_crossing\_human\_control 1.000000000  
## pedestrian\_crossing\_physical\_facilities 0.479999186  
## light\_conditions 0.008712157  
## weather\_conditions 0.179607849  
## road\_surface\_conditions 0.270071902  
## special\_conditions\_at\_site 0.474680204  
## carriageway\_hazards 0.465698657  
## urban\_or\_rural\_area -0.135655701  
## did\_police\_officer\_attend\_scene\_of\_accident 0.368834315  
## trunk\_road\_flag 0.066875310  
## severe\_accident -0.023200397  
## cluster 0.145122908  
## pedestrian\_crossing\_physical\_facilities  
## accident\_year NA  
## police\_force -0.214809833  
## accident\_severity 0.053215979  
## number\_of\_vehicles -0.045457006  
## number\_of\_casualties -0.063618862  
## day\_of\_week -0.001157246  
## local\_authority\_district NA  
## first\_road\_class -0.108747343  
## first\_road\_number 0.027107397  
## road\_type 0.022836398  
## speed\_limit -0.250591065  
## junction\_detail 0.263016395  
## junction\_control 0.172148522  
## second\_road\_class 0.160896869  
## second\_road\_number 0.072332373  
## pedestrian\_crossing\_human\_control 0.479999186  
## pedestrian\_crossing\_physical\_facilities 1.000000000  
## light\_conditions -0.003134603  
## weather\_conditions 0.108174516  
## road\_surface\_conditions 0.166821951  
## special\_conditions\_at\_site 0.300249069  
## carriageway\_hazards 0.295252075  
## urban\_or\_rural\_area -0.267252306  
## did\_police\_officer\_attend\_scene\_of\_accident 0.232236537  
## trunk\_road\_flag 0.093760011  
## severe\_accident -0.029113100  
## cluster 0.071175117  
## light\_conditions weather\_conditions  
## accident\_year NA NA  
## police\_force 0.0025391401 -0.071024180  
## accident\_severity -0.0520089420 0.041793950  
## number\_of\_vehicles -0.0657350778 -0.039831850  
## number\_of\_casualties 0.0375239899 -0.030356726  
## day\_of\_week 0.0049912472 -0.005787318  
## local\_authority\_district NA NA  
## first\_road\_class -0.0374431532 0.008133094  
## first\_road\_number 0.0041845919 -0.002755031  
## road\_type -0.0063354788 0.051925721  
## speed\_limit 0.0846792809 -0.041759182  
## junction\_detail -0.0034061279 0.102914596  
## junction\_control -0.0562853142 0.041029294  
## second\_road\_class -0.0578152012 0.017208955  
## second\_road\_number 0.0005423747 -0.001532152  
## pedestrian\_crossing\_human\_control 0.0087121568 0.179607849  
## pedestrian\_crossing\_physical\_facilities -0.0031346033 0.108174516  
## light\_conditions 1.0000000000 0.109051719  
## weather\_conditions 0.1090517188 1.000000000  
## road\_surface\_conditions 0.1543714072 0.362715245  
## special\_conditions\_at\_site 0.0213851839 0.184058028  
## carriageway\_hazards 0.0358522176 0.189291082  
## urban\_or\_rural\_area 0.0629603524 -0.037427688  
## did\_police\_officer\_attend\_scene\_of\_accident -0.0605649807 0.148632962  
## trunk\_road\_flag -0.0096957204 0.021753569  
## severe\_accident 0.0529544705 -0.013827202  
## cluster 0.0991988861 0.833764040  
## road\_surface\_conditions  
## accident\_year NA  
## police\_force -0.0500405958  
## accident\_severity 0.0188910768  
## number\_of\_vehicles -0.0300190386  
## number\_of\_casualties -0.0085895625  
## day\_of\_week -0.0046302108  
## local\_authority\_district NA  
## first\_road\_class -0.0140658248  
## first\_road\_number -0.0081124341  
## road\_type 0.0802982960  
## speed\_limit 0.0063155994  
## junction\_detail 0.1536932283  
## junction\_control 0.0641129132  
## second\_road\_class 0.0098480564  
## second\_road\_number -0.0072207121  
## pedestrian\_crossing\_human\_control 0.2700719018  
## pedestrian\_crossing\_physical\_facilities 0.1668219505  
## light\_conditions 0.1543714072  
## weather\_conditions 0.3627152449  
## road\_surface\_conditions 1.0000000000  
## special\_conditions\_at\_site 0.3334905594  
## carriageway\_hazards 0.3532930908  
## urban\_or\_rural\_area 0.0049411510  
## did\_police\_officer\_attend\_scene\_of\_accident 0.0854875044  
## trunk\_road\_flag -0.0076529688  
## severe\_accident -0.0002743842  
## cluster 0.3327792373  
## special\_conditions\_at\_site  
## accident\_year NA  
## police\_force -0.123744405  
## accident\_severity 0.041804294  
## number\_of\_vehicles -0.005585140  
## number\_of\_casualties -0.032978238  
## day\_of\_week -0.003216982  
## local\_authority\_district NA  
## first\_road\_class -0.020955881  
## first\_road\_number -0.023547258  
## road\_type 0.117024343  
## speed\_limit -0.066288267  
## junction\_detail 0.287368813  
## junction\_control 0.112068771  
## second\_road\_class 0.041169295  
## second\_road\_number -0.010710349  
## pedestrian\_crossing\_human\_control 0.474680204  
## pedestrian\_crossing\_physical\_facilities 0.300249069  
## light\_conditions 0.021385184  
## weather\_conditions 0.184058028  
## road\_surface\_conditions 0.333490559  
## special\_conditions\_at\_site 1.000000000  
## carriageway\_hazards 0.620653965  
## urban\_or\_rural\_area -0.053487850  
## did\_police\_officer\_attend\_scene\_of\_accident 0.222291104  
## trunk\_road\_flag 0.029670656  
## severe\_accident -0.016723587  
## cluster 0.154156717  
## carriageway\_hazards  
## accident\_year NA  
## police\_force -0.12113420  
## accident\_severity 0.03638516  
## number\_of\_vehicles -0.02149395  
## number\_of\_casualties -0.03624207  
## day\_of\_week -0.00242029  
## local\_authority\_district NA  
## first\_road\_class -0.01600646  
## first\_road\_number -0.02469392  
## road\_type 0.12256812  
## speed\_limit -0.06684770  
## junction\_detail 0.28023186  
## junction\_control 0.10727081  
## second\_road\_class 0.03332163  
## second\_road\_number -0.01193078  
## pedestrian\_crossing\_human\_control 0.46569866  
## pedestrian\_crossing\_physical\_facilities 0.29525207  
## light\_conditions 0.03585222  
## weather\_conditions 0.18929108  
## road\_surface\_conditions 0.35329309  
## special\_conditions\_at\_site 0.62065396  
## carriageway\_hazards 1.00000000  
## urban\_or\_rural\_area -0.05535304  
## did\_police\_officer\_attend\_scene\_of\_accident 0.19781814  
## trunk\_road\_flag 0.03130437  
## severe\_accident -0.01275810  
## cluster 0.16114644  
## urban\_or\_rural\_area  
## accident\_year NA  
## police\_force 0.295660891  
## accident\_severity -0.102508851  
## number\_of\_vehicles 0.042968365  
## number\_of\_casualties 0.144004033  
## day\_of\_week -0.005515087  
## local\_authority\_district NA  
## first\_road\_class -0.179821372  
## first\_road\_number 0.002215652  
## road\_type -0.048734379  
## speed\_limit 0.673260664  
## junction\_detail -0.131103860  
## junction\_control -0.226458176  
## second\_road\_class -0.264616413  
## second\_road\_number -0.022395946  
## pedestrian\_crossing\_human\_control -0.135655701  
## pedestrian\_crossing\_physical\_facilities -0.267252306  
## light\_conditions 0.062960352  
## weather\_conditions -0.037427688  
## road\_surface\_conditions 0.004941151  
## special\_conditions\_at\_site -0.053487850  
## carriageway\_hazards -0.055353037  
## urban\_or\_rural\_area 1.000000000  
## did\_police\_officer\_attend\_scene\_of\_accident -0.183882202  
## trunk\_road\_flag -0.192030442  
## severe\_accident 0.083277500  
## cluster 0.022584016  
## did\_police\_officer\_attend\_scene\_of\_accident  
## accident\_year NA  
## police\_force -0.182403388  
## accident\_severity 0.190422214  
## number\_of\_vehicles -0.010457309  
## number\_of\_casualties -0.126197117  
## day\_of\_week -0.001149979  
## local\_authority\_district NA  
## first\_road\_class 0.045718076  
## first\_road\_number -0.038372190  
## road\_type 0.055448642  
## speed\_limit -0.226355179  
## junction\_detail 0.257163580  
## junction\_control 0.127507315  
## second\_road\_class 0.076890111  
## second\_road\_number 0.002234487  
## pedestrian\_crossing\_human\_control 0.368834315  
## pedestrian\_crossing\_physical\_facilities 0.232236537  
## light\_conditions -0.060564981  
## weather\_conditions 0.148632962  
## road\_surface\_conditions 0.085487504  
## special\_conditions\_at\_site 0.222291104  
## carriageway\_hazards 0.197818136  
## urban\_or\_rural\_area -0.183882202  
## did\_police\_officer\_attend\_scene\_of\_accident 1.000000000  
## trunk\_road\_flag 0.124601813  
## severe\_accident -0.066216908  
## cluster 0.087604670  
## trunk\_road\_flag severe\_accident  
## accident\_year NA NA  
## police\_force -0.6210245233 0.0451418652  
## accident\_severity 0.0730154170 -0.4643089019  
## number\_of\_vehicles -0.0080377244 -0.0088722152  
## number\_of\_casualties -0.0581187467 0.0672717209  
## day\_of\_week 0.0006439579 0.0002434010  
## local\_authority\_district NA NA  
## first\_road\_class 0.1320770718 -0.0337179581  
## first\_road\_number -0.0159725741 0.0078624445  
## road\_type 0.0623080734 0.0006919176  
## speed\_limit -0.2461658592 0.0961383526  
## junction\_detail 0.0611102073 -0.0244314541  
## junction\_control 0.1082377347 -0.0555999567  
## second\_road\_class 0.1181246856 -0.0530937964  
## second\_road\_number 0.0014039466 -0.0148863514  
## pedestrian\_crossing\_human\_control 0.0668753096 -0.0232003966  
## pedestrian\_crossing\_physical\_facilities 0.0937600113 -0.0291131003  
## light\_conditions -0.0096957204 0.0529544705  
## weather\_conditions 0.0217535689 -0.0138272023  
## road\_surface\_conditions -0.0076529688 -0.0002743842  
## special\_conditions\_at\_site 0.0296706562 -0.0167235867  
## carriageway\_hazards 0.0313043677 -0.0127580958  
## urban\_or\_rural\_area -0.1920304424 0.0832775004  
## did\_police\_officer\_attend\_scene\_of\_accident 0.1246018126 -0.0662169078  
## trunk\_road\_flag 1.0000000000 -0.0480466238  
## severe\_accident -0.0480466238 1.0000000000  
## cluster -0.0047876409 0.0130523731  
## cluster  
## accident\_year NA  
## police\_force -0.042038631  
## accident\_severity 0.003293713  
## number\_of\_vehicles 0.134210549  
## number\_of\_casualties 0.307315245  
## day\_of\_week -0.001440695  
## local\_authority\_district NA  
## first\_road\_class -0.032184856  
## first\_road\_number -0.002904445  
## road\_type 0.034422774  
## speed\_limit 0.036808507  
## junction\_detail 0.080967892  
## junction\_control 0.008917945  
## second\_road\_class -0.012586111  
## second\_road\_number -0.006114766  
## pedestrian\_crossing\_human\_control 0.145122908  
## pedestrian\_crossing\_physical\_facilities 0.071175117  
## light\_conditions 0.099198886  
## weather\_conditions 0.833764040  
## road\_surface\_conditions 0.332779237  
## special\_conditions\_at\_site 0.154156717  
## carriageway\_hazards 0.161146441  
## urban\_or\_rural\_area 0.022584016  
## did\_police\_officer\_attend\_scene\_of\_accident 0.087604670  
## trunk\_road\_flag -0.004787641  
## severe\_accident 0.013052373  
## cluster 1.000000000