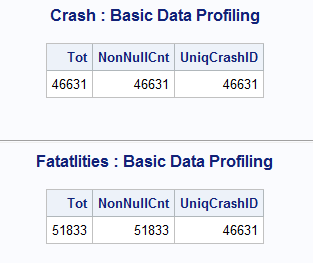
Question 1

Data Profiling of Crash and Fatalities



The Crash data has all the unique crashes while fatalities has multiple records per crashes.

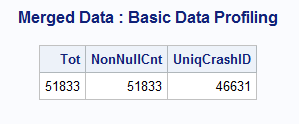
Relationship Description

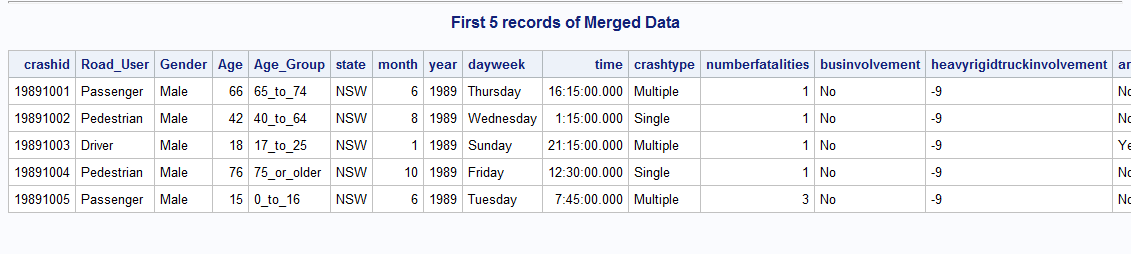
The fatalities dataset has multiple records or casualties for some crashes.

This is going to be One-to-Many relationship. One Crash getting mapped to Multiple casualties.

Make sure there are no duplicates, which i have checked above

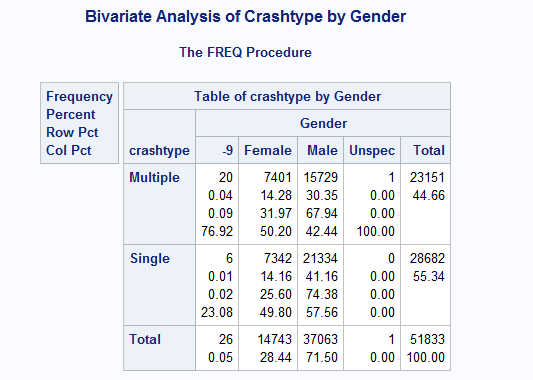
Since the attribute crashid has different names in each dataset, it needs to be renamed and both the datasets need to be sorted before merging.

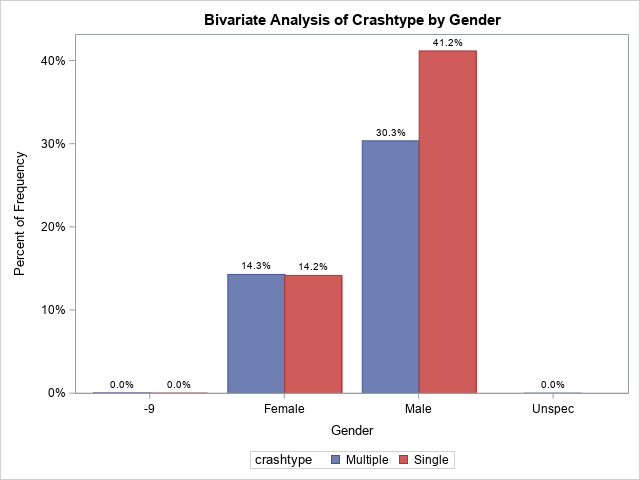




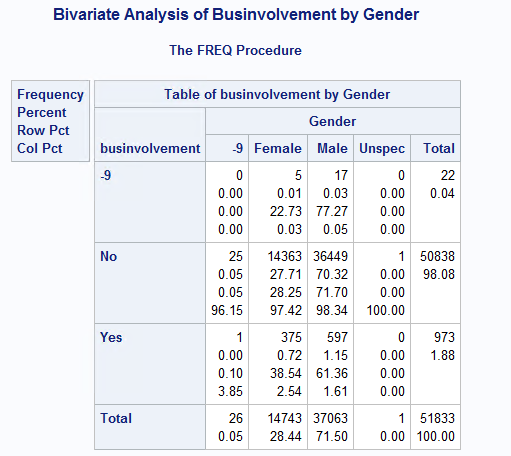
The merge is implemented correctly.

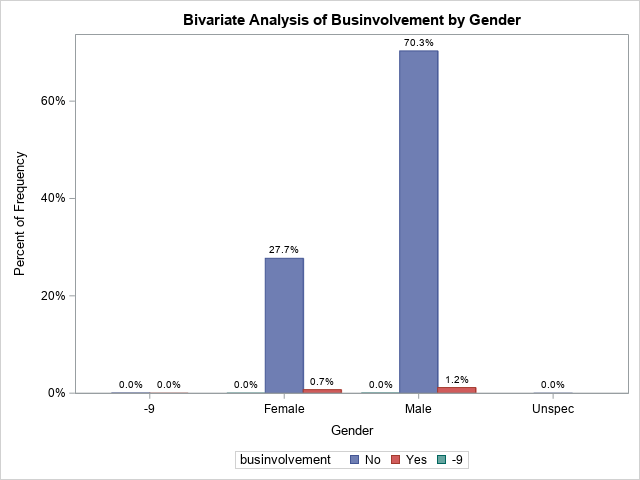
Question 2



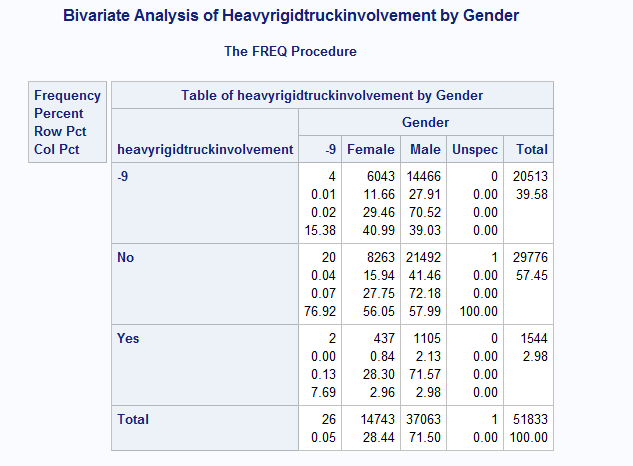


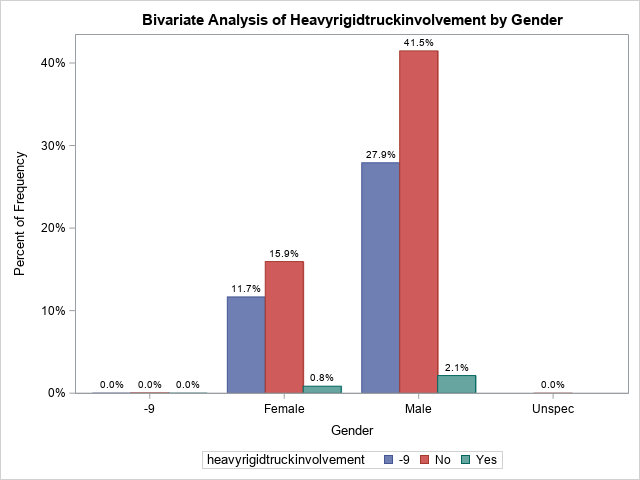
We can infer that Males tend to indulge in more Single accident.





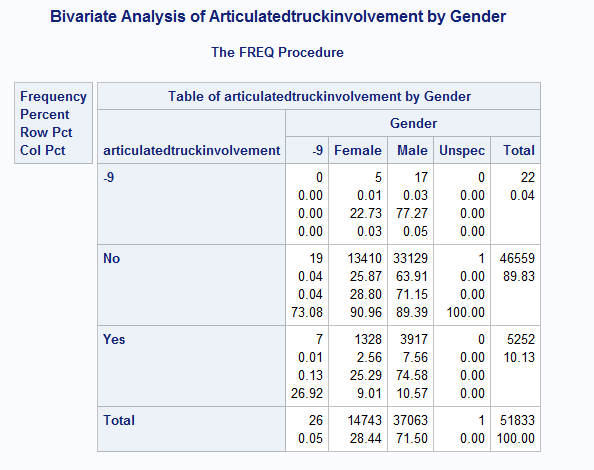
Very few crashes have bus involvement irrespective of gender.

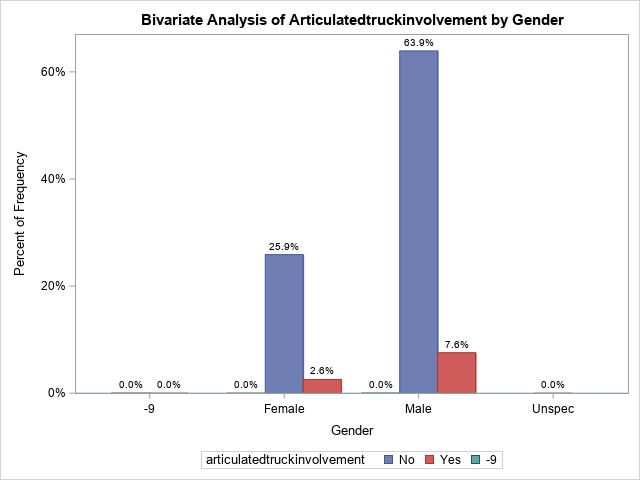


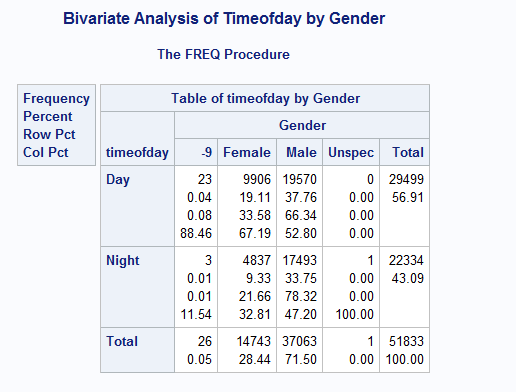


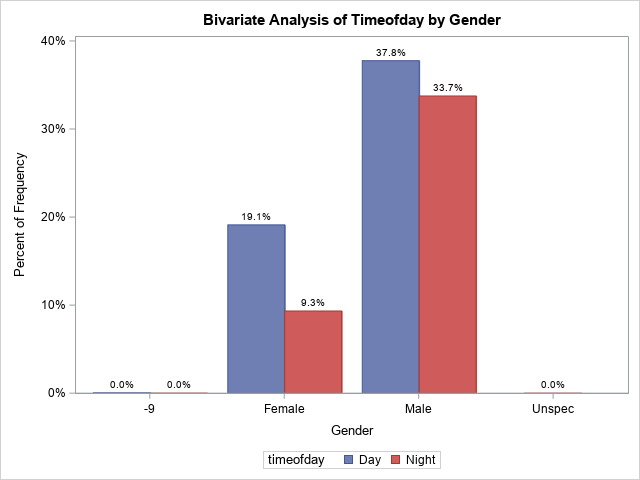
Heavy truck involvement in crash is less and there are many crashes without this information.

Approximately 2% of accidents involving male is with heavy truck, which is more than female.

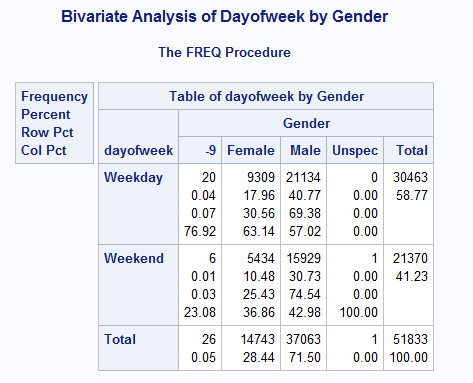


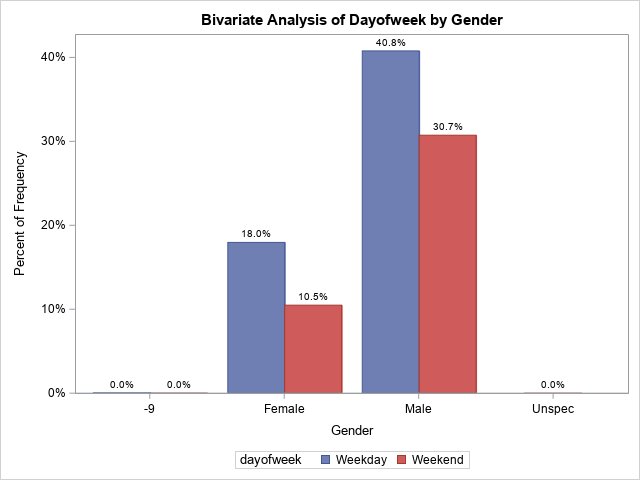






Female tend to have more accidents during Daytime. One of the reason maybe Because Female avoid travelling at Night time. And hence there is a bigger gap between day and night for female in compared to male, who have almost similar rate across day and night.





It is observed more crashes happen in weekday as it constitutes 5 days as compared to 2 days of weekends.

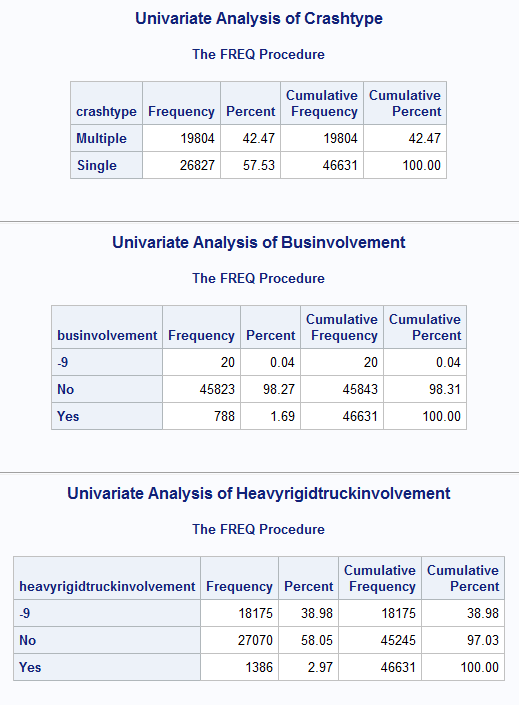
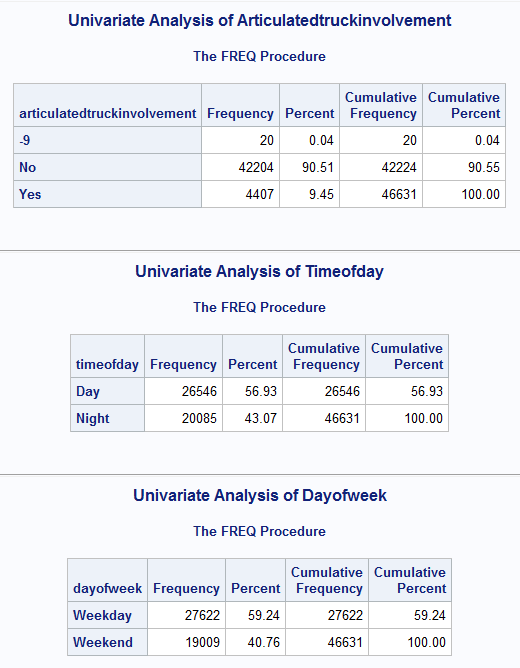
Another reason is people tend to stay at home in contrast to the office/work travelling in weekdays.

It applies for both the gender.

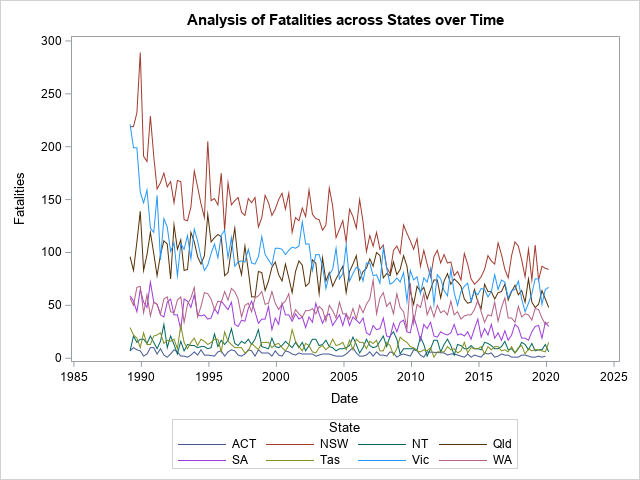
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Number of crashes** | | **Number of fatalities** | | |  |
|  |  | **Total**  **(n= )** | **Percent**  **\*** | **Total**  **(n= )** | **Females**  **(n= )** | **Males**  **(n= )** | **% Male** |
| Crash type | Multiple | 19804 | 42 | 23151 | 7401 | 15729 | 30 |
|  | Single | 26827 | 58 | 28682 | 7342 | 21334 | 41 |
| Bus  involved | No | 45823 | 98 | 50838 | 14363 | 36449 | 70 |
|  | Yes | 788 | 1.7 | 973 | 375 | 597 | 1.2 |
| Heavy rigid truck  involved | No | 27070 | 58 | 29776 | 8263 | 21492 | 41 |
|  | Yes | 1386 | 3.0 | 1544 | 437 | 1105 | 2.1 |
| Articulated truck  involved | No | 42204 | 90 | 46559 | 13410 | 33129 | 64 |
|  | Yes | 4407 | 9.5 | 5252 | 1328 | 3917 | 7.6 |
| Time of day | Day | 26546 | 57 | 29499 | 9906 | 19570 | 38 |
|  | Night | 20085 | 43 | 22334 | 4837 | 17493 | 34 |
| Time of  week | Weekday | 27622 | 59 | 30463 | 9309 | 21134 | 41 |
|  | Weekend | 19009 | 41 | 21370 | 5434 | 15929 | 31 |

\*percentages that don’t add up to 100% are due to missing/unknown data.

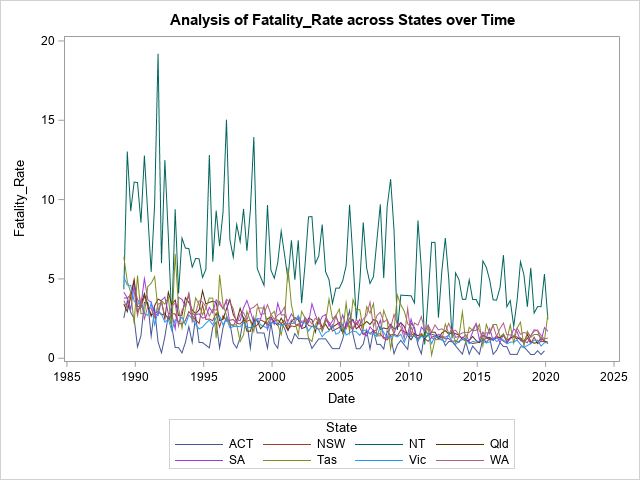
The numbers for first part of the table are obtained from these results by PROC FREQ. and for second part the above charts are used.

Question 3



The Crashes has decreased over time and is clearly visible from this time series graph. But there is a huge drop for New South Whales, Victoria and Queensland, while for other the count has been more of constant.



The fatality rate has decreased drastically for NT, while there is a consistent drop for other states also over these years.

These are individual plots of Fatality Rate across states over time and In All the states it has decreased.

