MCBI71: Business Intelligence Certificate to Support Managerial Decisions

Australian road accidents

Assessment 3

A group of people in a convertible car

Description automatically generated

Rojesh Dhakal

December 17, 2023

Table of Content

[1. Introduction 3](#_Toc153463407)

[2. Review of EDA 4](#_Toc153463408)

[2.1. Review of Assessments 1 4](#_Toc153463409)

[2.2. Review of Assessment 2 5](#_Toc153463410)

[3. Statistical Inference 7](#_Toc153463411)

[3.1 Dataset Overview 8](#_Toc153463412)

[3.2 Variable Types 8](#_Toc153463413)

[3.3 Cross Tabulations 9](#_Toc153463414)

[3.3.1 Gender and Age group 9](#_Toc153463415)

[3.3.2 Gender & Time of Day 9](#_Toc153463416)

[3.4 Hypothesis Testing 10](#_Toc153463417)

[3.4.1 Analysis of road accidents with speed limit 10](#_Toc153463418)

[3.4.2 Analysis of crash with years 10](#_Toc153463419)

[3.4.3 Analysis between crash with age 11](#_Toc153463420)

[3.4.4 Analysis between age and speed 11](#_Toc153463421)

[3.5 Time Series Analysis 12](#_Toc153463422)

[3.5.1 Naïve Forecast 12](#_Toc153463423)

[3.5.2 Time Series Analysis with linear trend 13](#_Toc153463424)

[3.5.3 Holt’s Smoothening 14](#_Toc153463425)

[4. Results Evaluation 16](#_Toc153463426)

[4.1 Findings from Cross Tabulation 16](#_Toc153463427)

[4.2 Findings from Time Series Analysis 16](#_Toc153463428)

[5. Insights 17](#_Toc153463429)

[6. Conclusions 18](#_Toc153463430)

[7. References 19](#_Toc153463431)

[8. Appendix 20](#_Toc153463432)

**List of Figures**

[Figure i: Dashboard 5](#_Toc153445414)

[Figure ii: Crash by different Age-group and Speed 6](file:///C:\Users\User\Desktop\Future%20Skills\Assignment%203\Template-A3(MCBI71)%20(1).docx#_Toc153445415)

[Figure iii: Crash by Gender and Speed 7](file:///C:\Users\User\Desktop\Future%20Skills\Assignment%203\Template-A3(MCBI71)%20(1).docx#_Toc153445416)

[Figure iv: Crash by Gender and Age-group during Day/Night 7](file:///C:\Users\User\Desktop\Future%20Skills\Assignment%203\Template-A3(MCBI71)%20(1).docx#_Toc153445417)

[Figure v: Cross tabulation between Gender and Age-group 9](#_Toc153445418)

[Figure vi: Cross tabulation between Gender & Time of Day 10](#_Toc153445419)

[Figure vii: Naive Forecasting 13](#_Toc153445420)

[Figure viii: Forecasting using LinearTrend 14](#_Toc153445421)

[Figure ix: Holt's Smoothing Forecast 15](#_Toc153445422)

[Figure x: Forecast Sheet 16](#_Toc153445423)

[Figure xi: Crash timings of Young Adults 19](#_Toc153445424)

[Figure xii: Involvement in crash by different Age-groups at the Speed of 100 20](#_Toc153445425)

**List of Tables**

[Table 1: Hypothesis testing between Crash ID & Speed Limit 11](#_Toc153445495)

[Table 2: Hypothesis testing between Crash ID and Year 11](#_Toc153445496)

[Table 3: Hypothesis testing between Crash ID & Age 11](#_Toc153445497)

[Table 4: Hypothesis testing between Speed Limit & Age 12](#_Toc153445498)

[Table 5: Metrics Evaluation 17](#_Toc153445499)

# Introduction

Road traffic accidents stands out as a major cause of both fatalities and disabilities worldwide, which leads to substantial economic burden. Notably, it is anticipated that by the year 2030, road traffic accidents will rank as the fifth leading cause of death globally (Mannering & Bhat, 2014). This report offers a thorough analysis of Australian Road Accident data spanning from 1989 to 2021. The primary focus lies in looking deep into behaviours, relationships, and emerging trends through the utilization of cross-tabulation, time series analysis and hypothesis testing. The assessment aims to provide valuable insights that can inform strategic decisions, improve road safety measures, and contribute to road safety.

When Connelly and Supangan (2006), calculated the total economic costs incurred by the road accidents in Australia, they found that the total accidents accounted for the 2.3% of the gross domestic product annually. These costs ranged between 0.62% and 3.63% of the gross state product. In certain instances, costs related to road accidents can be categorized as either direct or indirect. Direct costs involve expenses incurred by public health services, while indirect costs adversely affect the national impact through various channels. These indirect costs stem from the loss of national income due to decreased productivity and the reduction of labour hours among the workforces. (Damien,1994; Jagnoor et al., 2015). According to Gorea (2016), the financial repercussions of road accidents have both immediate and lasting consequences on victims and their families. This report seeks to comprehensively grasp the main causes of crash accidents through power of data visualization, inferential statistics, and time series analysis. The primary focus lies in creating safety measures and policies, which will help in reducing tragic incidents and save lives which ultimately saves the economy of the country.

The primary objective of this report is to conduct a comprehensive time series analysis on the Australian road accidents datasets. Through this analysis, the goal is to extract valuable insights and information, specially focusing in on various factors, patterns, and trends that exert a significant influence on the crash fatalities over different time periods. This comprehensive examination has been further enhanced by evaluating dataset using Naïve, Linear trend and Holt’s smoothing methods. The utilization of these techniques aims to provide a more nuanced understanding of the temporal dynamics of road accidents and the development of targeted interventions to enhance overall road safety. A new business question arises: What are the long-term trends in road accidents over the entire time span (1989 to 2021) and what significant shifts or patterns can be uncovered by employing time series analysis?

To achieve this objective, a comprehensive analysis of the dataset has been undertaken, employing cross tabulation and time series techniques alongside inferential statistics. Through meticulous data preparation and exploration, critical factors contributing to crashes have been identified over different time periods. The integration of time series analysis allows for a nuanced understanding of temporal trends, while cross-tabulation facilitates the exploration of relationships among variables. Leveraging statistical inference, evidence-based recommendations have been formulated to inform safety policies and measures. These strategies are designed to not only address the identified factors but also to adapt to evolving trends highlighted by time series analysis, ultimately aiming to reduce the tragic toll of lives lost in the crashes.

# Review of EDA

## Review of Assessments 1

In the initial assessment, the primary objective was to develop and implement effective road safety measures which can help to reduce the number of fatalities and improve overall safety. This phase involved a meticulous process of cleaning and transforming the data to ensure its integrity and reliability. The resulting report provided fundamental insights into the nature of crashes, the prevalence of crashes such as in day or night, the age group that were prone to crashes, crashes occurrence during the holiday season, and total male crashes. The dashboard below provides a comprehensive overview of crashes, which has been derived with the help of analysis.

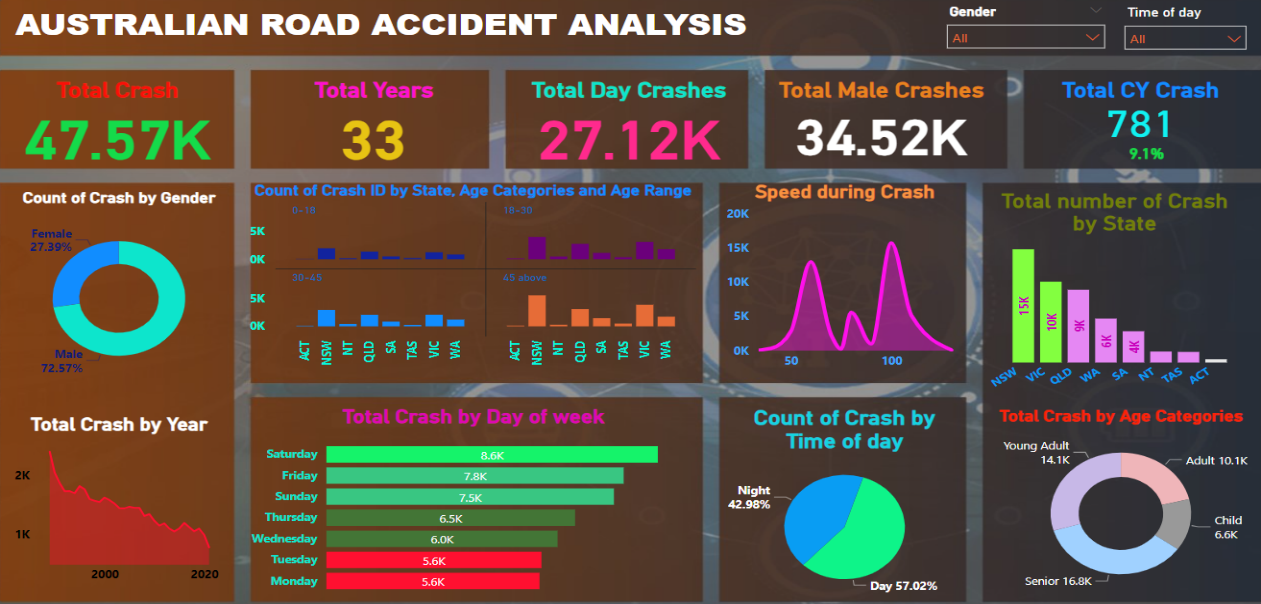


Figure : Dashboard

The above dashboard provides a compelling story, revealing about 47.57 thousand crashes occurred in Australia over a span of 33 years. Male were involved in a crash more which was about 72 percent accounting for 34.52 thousand crashes. Day time accidents were more prevalent, and Saturday exhibits a higher number of crash incidences. Senior populations were involved in a crash and considered a vulnerable group. NSW had a greater number of crash incidents as compared to another state. These insights contributed to understand the overall landscape of road accidents and helped to lay foundation for targeted interventions and policy consideration to enhance road safety across Australia. This analysis also uncovered pivotal information that informed and guided further for ongoing assessments.

## Review of Assessment 2

In the second assessment, the dataset has been explored in depth by a combination of descriptive and diagnostics analysis which sheds light on various factors, patterns, and trends which played a vital role in accidents. Leveraging inferential statistics and data visualization, the vulnerable groups have been identified, thereby laying the groundwork for targeted interventions to address their needs so that road safety and policy can be improved. Specific tailored strategies and campaigns can be developed according to the age group which can be more effective to raise awareness and encourage safer road behaviors. Some of the key information obtained from the analysis has been discussed below.

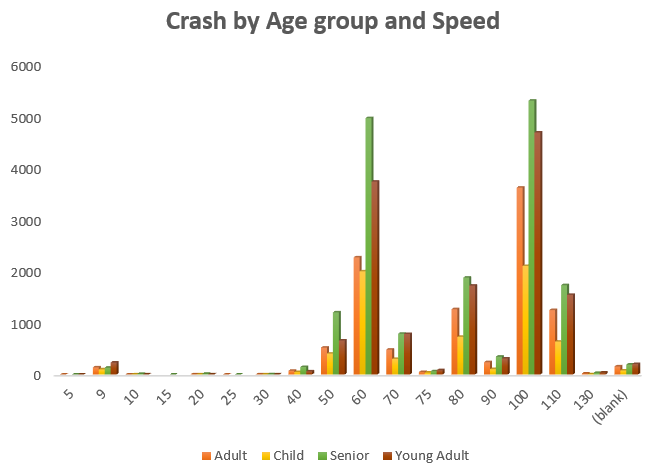
 The visual representation above paints a clear picture of accident distribution, higlighting the maximum number of accidents occurred at speed of 100, where senior population were more likely to be involved in a crash closely followed by young adults. The second most hazardous speed limit, where signicficant number of accidents occurred, was at 60, where senior again stood high followed by young adults. It is noteworthy that senior were considered as the top contributors to the accidents in various speed limits, with young adults following them as the next significant group. Further analysis and awareness campaigns may be benefical to enhance road safety for both seniors and young adults.

Figure ii: Crash by different Age-group and Speed

A graph of a graph with numbers and text

Description automatically generated with medium confidence The above figure unfolds a compelling narrative, which presents a notable disparity in gender wise crash occurences across all speed limits. Overall, the male population experienced a higher number of crash incidents. The detailed analysis presents that crashes were more prevalent at the speed of 100, where female populations accounted for 4.3 thousand incidents. However, the male poulation significantly outpaced this number, which accounted for 11.3 thousand incidents- which was more than double the count for females. Examining the second-highest crash occurrence speed, which was at 60, the gender disparity remains evident. Males accounted for 9.2 thousand incidents, whereas females on the other hand accounted for one-third of the number, which is 3.6 thousand crashes. This data underscores a clear trends in gender disparity concerning crash rates across various speed limits, emphasizing the need for targeted interventions and awareness campaigns.

Figure iii: Crash by Gender and Speed

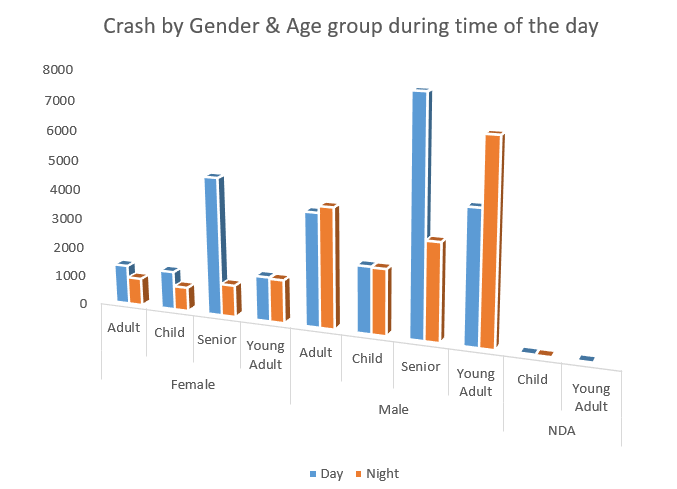


Figure iv: Crash by Gender and Age-group during Day/Night

The provided data offers crucial insights of crashes which is categorized by various genders, age groups, and time of day. A notable distinction is evident between day and night, with 20.4 thousand crashes occurring during night and 27.1 thousand crashes occurring during day. Notably, seniors in both genders, significantly contributed to crashes, emphasizing the need for targeted safety measures for this demographic. It’s worth highlighting male young adults had a higher involvement in crashes during the night as compared to other age group. This suggests a potential link to specific lifestyle or behaviour patterns, highlighting the importance of further analysis. Developing strategies to address and mitigate these risks could contribute to overall road safety.

# Statistical Inference

Statistical inference, as outline by Makar and Rubin (2018), serves as a foundation of statistics which enables to formulate meaningful and evidence-based claims in situations where only incomplete data are available. This process is essential for drawing conclusions in the presence of uncertainty. Informal statistical inference is the process of making conclusions beyond the available data using three main principles: making generalization, treating data as evidence, and employing informal probability. These involve making claims that extend beyond the collected data, and the associated uncertainty is expressed through informal probability assessments of the likelihood of those claims (Makar & Rubin,2009). In this report statistical inference, plays a crucial role in hypothesis testing, where research questions are formulated, and through statistical tests, the assessments are done to check whether observed patterns or differences in the samples are significant. Beyond that, it helped in making predictions about future observations or trends within the population. Ultimately, this process contributes to the foundation of evidence-based decision making.

* 1. Dataset Overview

The dataset is in csv format and has been collected by the author directly from Australian Road Death Database and can be downloaded from Kaggle. There is over 3 decades of road accident data which ranges from 1989 to 2021. The dataset serves as a valuable resource for conducing in depth analysis and gaining insights of road accidents in Australia. The dataset can be accessed using this link <https://www.kaggle.com/datasets/deepcontractor/australian-fatal-car-accident-data-19892021>

* 1. Variable Types

|  |  |  |
| --- | --- | --- |
| Variable Category | Variable | Data Type |
| Unique Identifier | Crash Id | Categorical, Nominal |
| City | State | Categorical |
| Time Information | Month | Categorical, Ordinal |
| Year | Discrete, Continuous |
| Day Week | Categorical, Nominal |
| Time | Continuous |
| Day of week | Ordinal, Categorical |
| Time of day | Nominal, Categorical |
| Limits | Speed limit | Ordinal |
| Demographics | Age | Continuous |
| Age group | Categorical |
| Age categories | Categories, Ordinal |
| Road User | Categorical |
| Type | Crash Type | Categorical |
| User | Road User | Categorical |
| Roads | National Road Type | Nominal, Categorical |
| National Remoteness Area | Nominal, Categorical |

* 1. Cross Tabulations

### Gender and Age group

A screenshot of a child service

Description automatically generated A graph of different colored bars

Description automatically generated

Figure : Cross tabulation between Gender and Age-group

The cross-tabulation analysis between gender and age group reveals insightful patterns within the dataset. The distribution of gender demonstrates the significant difference in crash accidents, where male accounted for 34.5 thousand incidences whereas female only accounted for 13 thousand crashes. Regarding the age group, senior population were most likely to be involved in the crashes followed closely by young adults. Looking closely, young adult males recorded the highest number of crashes overall, totalling for 11.12 thousand incidents, closely followed by senior male at 11 thousand. Among females, senior population stood out with 5.7 thousand incidents while, young adult females accounted 2.9 thousand incidences. The cross tabulation provided a comprehensive view of interplay between gender and age categories, offering valuable insights and potential areas for more in-depth investigations.

### Gender & Time of Day

A screenshot of a computer

Description automatically generated A graph of a bar graph

Description automatically generated with medium confidence

Figure : Cross tabulation between Gender & Time of Day

The cross-tabulation analysis between gender and time of day reveals insightful patterns within the dataset. The distribution of gender demonstrates the significant difference in crash accidents, where male accounted for 34.5 thousand incidences whereas female only accounted 13 thousand crashes. A comprehensive overview of the time-of-day, analysis reveals that daytime crashes were more prevalent, which accounted for about 27 thousand crashes. During daytime, male was involved in 18.3 thousand crashes, whereas female only had 8.7 thousand crashes. Surprisingly during nighttime, female had 4.2 thousand crashes while male had four time more, which totalled for 16.18 thousand crashes. The cross tabulation provided a comprehensive view of interplay between gender and day of time, offering valuable insights and potential areas for more in-depth investigations into the factors influencing these patterns.

## Hypothesis Testing

### Analysis of road accidents with speed limit

Null hypothesis (H0): The relationship between speed limits and the occurrence of road accidents is not statistically significant.

Alternate hypothesis (H1): There exists a statistically significant association between speed limits and the frequency of road accidents.

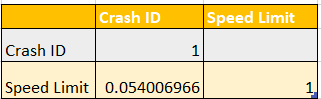


Table : Hypothesis testing between Crash ID & Speed Limit

The obtained value from the analysis is 0.054, slightly exceeding the predetermined significance level (alpha) of 0.05. As a result, the statistical inference leads to the non-rejection of the null hypothesis. This outcome implies that there is insufficient evidence to assert a statically significant relationship between speed limits and accident counts.

### Analysis of crash with years

Null hypothesis (H0): The incidence of accident crashes remains consistent across the years.

Alternate hypothesis (H1): There are fluctuations in the number of accident crashes over the years.

A white rectangular sign with black text

Description automatically generated

Table : Hypothesis testing between Crash ID and Year

The obtained value from analysis is 0.99 which surpasses the significance level of 0.05. The high p-value indicates a lack of statistical evidence to reject null hypothesis, which speculates that the number of accidents remains constant over time. Consequently, the alternate hypothesis, is not supported by the current analysis.

### Analysis between crash with age

Null hypothesis (H0): There is no substantial correlation between age and the occurrence of road accidents.

Alternate hypothesis (H1): A notable correlation exists between age and road accidents.

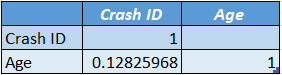


Table : Hypothesis testing between Crash ID & Age

The p-value obtained from the analysis of the correlation between age and road accidents is 0.12, which exceeds the conventional significance level of 0.05. The higher p-value suggests that there is insufficient statistical evidence to reject the null hypothesis, which suggests no substantial correlation between age and the occurrence of road accidents. Therefore, it cannot be confidently concluded that age is significantly corelated with the road accidents.

### Analysis between age and speed

Null hypothesis (H0): There is no substantial correlation between age and the speed of the vehicles involved in crashes.

Alternate hypothesis (H1): Age exhibits a correlation with the speed of vehicles involved in crashes.

A speed limit label with numbers

Description automatically generated

Table : Hypothesis testing between Speed Limit & Age

The p-value obtained from the analysis of the correlation between age and crash speed is -0.05, indicating a very weak negative correlation between these variables. This implies that, on average, as age increases, there is slight tendency for crash speed to decrease. Null hypothesis can be rejected, and it suggests that relationships is so weak that it has any substantial predictive or explanatory power.

## Time Series Analysis

A time series refers to set of recorded observations of clearly defined data acquired through repeated measurements conducted over a specific period (Australian Bureau of Statistics, 200). For many years, the field of road traffic accident analysis has heavily relied on statistical analysis and modelling to extract valuable insights from data. Professionals in this domain encourage for the application of statistical modelling, analysis, and forecasting to pinpoint the underlying causes of issues to lay the foundation for development of polices and interventions based on economic conditions (Zhang, Pang, Stallones, & Xiang, 2015). The integration of time series analysis into this framework proves invaluable, allowing for explorations of temporal trends, patterns, and recurring factors in road traffic.

### Naïve Forecast

D. Hyndman and H. Athanasopoulos (2016) argue that the naïve forecast, stands out as a straightforward forecasting technique despite its simplicity, is frequently identified as remarkably effective, often serving as the benchmark method for comparing various models. This forecasting assumes that future values of time series will be same as the most recent observed value.

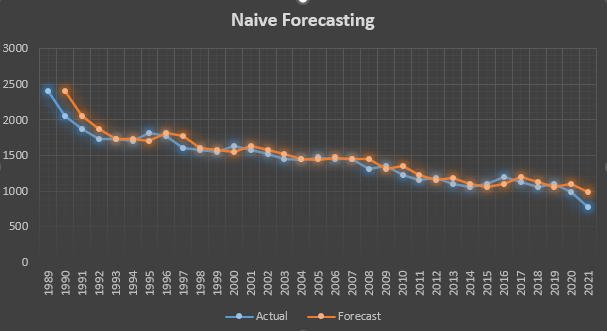


Figure : Naive Forecasting

The Mean Absolute Error (MAE) of 81.37 and the Root Mean Squared Error (RMSE) of 108.47 provide valuable metrics for evaluating the accuracy and precision of predictive model. MAE of 81.37 suggests a moderate level of accuracy in the predictions. With an RMSE of 108.47, there is wider dispersion of errors, which implies there is some variability in its predictions.

### Time Series Analysis with linear trend

After conducting the time series analysis, it presents a linear trend characterized by the equation -33\*x+2013.7. This model implies a consistent decrease over time, as indicated by the negative slope coefficient of -33x. The intercept, 2013.7, represents the expected value of dependent variable at the beginning of time series. The high R-squared value of 90.08 suggests that approximately 90.08% of the observed variability in the dependent variable can be attributed to the linear trend in time. Such a strong fit indicates that the linear model effectively captures the underlying pattern in the time series data. The insight forecasted for 2023 the crash number would be 835, which provides valuable information for policymakers, enabling them to anticipate trends and formulate targeted interventions for enhanced road safety measures.

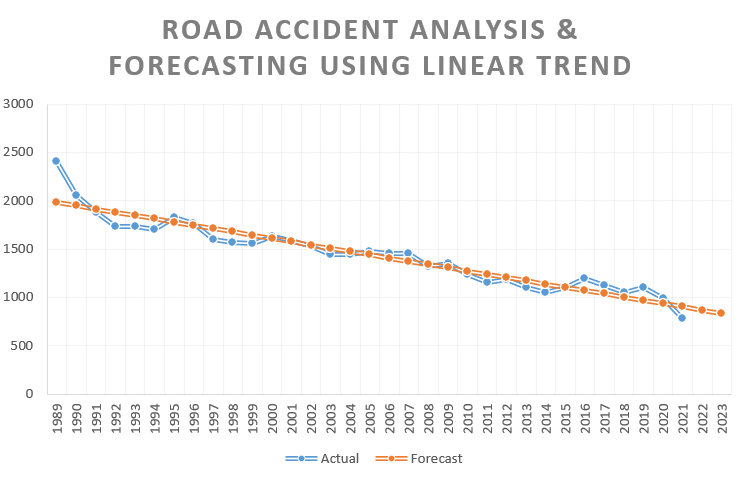


Figure : Forecasting using Linear Trend

The Mean Absolute Error (MAE) of 77.04 and the Root Mean Squared Error (RMSE) of 106.35 provide valuable metrics for evaluating the accuracy and precision of predictive model. MAE of 77.04 suggests a moderate level of accuracy in the predictions. With an RMSE of 106.35, there is wider dispersion of errors, which implies there is some variability in its predictions.

### Holt’s Smoothening

According to Mgale et al. (2021), predictions in seasonal times depends on three factors: the trend, level, and seasonal coefficient. Exponential smoothing is a technique to smooth time series, assigns exponentially diminishing weights to historical data, reducing the influence of past data. The chosen parameters (alpha=0.6, beta=0.5) were determined through iterative testing to optimize forecasting accuracy.

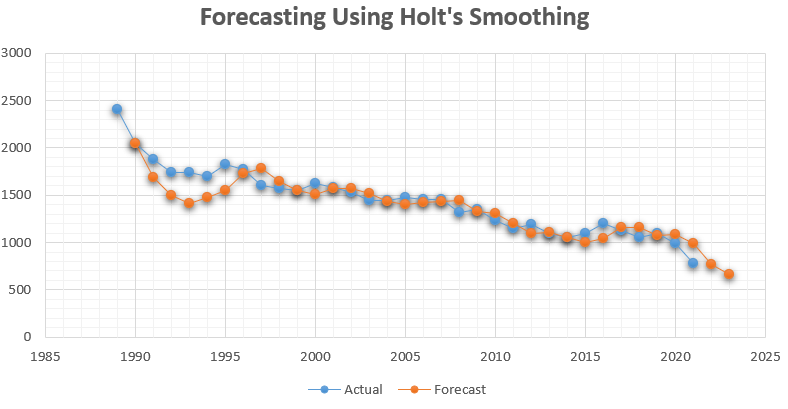


Figure : Holt's Smoothing Forecast

The calculated Mean Absolute Error (MAE) of 164.94 and Root Mean Squared Error (RMSE) of 437.64 in the analysis provide crucial metrics for assessing the accuracy of model. MAE of 164 indicates the average absolute difference between predicted and actual values, while the RMSE of 437, being a larger measure, underscores the spread and variability of these differences. Further investigation can be done to determine the cause of large errors and potential model refinements to improve the accuracy of future predictions.

* 1. **Forecast Sheet**

The data has also been forecasted using built-in function of Excel, which has been designed to simplify the process of creating forecasts or predictions based on historical data. It uses built-in forecasting functions to analyse trends and generate future forecasts.

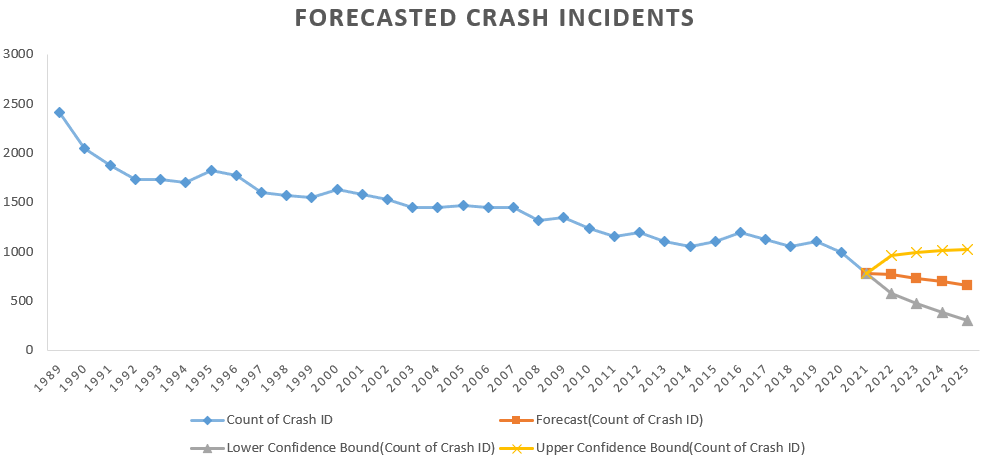


Figure : Forecast Sheet

The figure above shows the predicted number of road accidents using forecast-sheet. The central estimate, depicted as orange data point, signifies the predicted number of crashes for the upcoming four years. This prediction serves as the focal point or best estimate for the future. The lower confidence bound, represented by lower limit of grey region, delineates the minimum range within which the true crash count is anticipated to fall. The upper confidence bound is the upper end of the range within which the true crash count is expected to be, with certain level of confidence. For instance, in 2022, the forecasted value for road crashes is 765, with upper limit of 957, suggesting the highest expected crash count, and lower limit of 573, representing the lower end of the predicted range. These bounds offer valuable information about the range of uncertainty associated with the forecasted values.

# Results Evaluation

## Findings from Cross Tabulation

The findings of cross tabulation revealed that male population were prone to accidents, making 73% of the incidents. Among different age categories, senior males had the highest incident rate, followed by young adults. Daytime accidents were more prevalent, with male involved in 18.3 thousand accidents, while nighttime accidents totalled around 16 thousand crashes while female accounted for one fourth of the nighttime crashes which was 4 thousand.

## Findings from Time Series Analysis

Time series analysis and forecasting were conducted using three different models: Naïve Forecast, Linear Trend and Holt’s Smoothing. The evaluation of these models was based on performance metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results of this evaluation are summarized in the table below.

|  |  |  |
| --- | --- | --- |
|  | Mean Absolute Error | Root Mean Squared Error |
| Naïve Forecast | 81.37 | 108.47 |
| Linear Trend | 77.04 | 106.35 |
| Holt’s Smoothing | 164.94 | 437.64 |

Table : Metrics Evaluation

The table provides a comprehensive overview of the accuracy and effectiveness of each forecasting technique. Notably, Linear trend forecasting method emerged as the superior performer among the three models in terms of accuracy. With a lower MAE of 77.04, signifying the average absolute difference between the predicted and actual values, the Linear Trend model revealed a high precision in its predictions. Additionally, the RMSE for linear trend was 106.35, which is lower than other models, which indicates that model has smaller overall errors and is better at capturing variability in the data.

# Insights

The in-depth analysis outlined in Appendix 8.1 provides a vivid overview of the crash timings among young adults. A notable analysis reveals that, young adults were more involved in crash during nighttime, particularly between 10pm-12am. This analysis presents us an interesting trend and intriguing question about their social habits, lifestyle choice or potential visual impairments of young adults during late hours. Further research and investigation can be done into contributing factors and targeted interventions can be developed. Furthermore, police patrols can be increased targeting young adults to mitigate crash rates and promote safe driving practices. Interestingly, the pattern for crashes involving young females remains relatively steady, with the majority occurring from 4pm to 6 pm. The peak in accidents during this timeframe may be attributed to rush hour congestion and work-related activities, offering valuable insights into the dynamics of accidents during this timeframe.

Building upon previous analysis, it is evident that crashes occurring at speed of 100 pose a significant concern, which constitutes nearly one third of the total accidents. This underscores the pressing need for comprehensive safety measures. It is evident from Appendix 8.2, that senior and young adults are prominently involved in these high-speed crashes, which highlights the importance of effective awareness programs and education, targeted young driver training, and data-driven interventions. Additionally, further investigations are recommended to understand the specific locations and reasons behind the prevalence of 100 speed crashes. This can lead to adjustments in speed limits, the enforcement of traffic laws, and enhanced safety measures in identified high-risk zones.

The linear trend analysis reveals a decreasing trajectory in the number of accidents expected in coming years. The chart shows R squared value of 90.08 which indicates that it is the best fitted to the dataset, reinforcing the reliability of the linear trend model. The equation derived from the analysis (y=-33\*x+2013.7) helps policymakers to predict the number of accidents for a given year. This insight can guide proactive measures and interventions to address road safety concerns based on projected trends.

# Conclusions

The integration of time series analysis, incorporating Naïve method, linear trends, and Holt’s Smoothing techniques, has significantly provide valuable insights into the dynamics of road safety, effectively addressing the formulated business questions. Linear trend had the lowest Mean Absolute Error & Root Mean Squared Error which attests to the model’s accuracy in predicting trends over time. The application of these techniques, along with cross-tabulation and hypothesis testing, has allowed for comprehensive examination of the data. This comprehensive approach not only identified temporal patterns and fluctuations in road safety but also facilitated the formulation of targeted interventions. The findings have sufficiently addressed the targeted demographic groups, proposed age-specific measures and highlighting the significance of temporal considerations in road safety interventions. Additionally, speeding was major issues during the crashes, most accident happening at the speed of 100, which will help policymakers to design a new speed limit and further investigates why the crashes were occurring.

There are some limitations which needs to be acknowledged in the analysis, such as certain external factors influencing road safety which has not been fully accounted for. Also, linear trend model demonstrated lower MAE, the limitations lie in its assumption of a consistent linear pattern, which might oversimplify the complex dynamics of road safety. Therefore, future datasets need to be tested with other similar models before deploying it. Similarly, for Holt’s Smoothing using solver to find optimum value for alpha and beta. Additionally, a deeper investigation into external factors that contributed to crashes, such as weather conditions and infrastructure, accident severity could enhance the robustness of future assessments. Despite, these limitations, the insights gained from the analysis paved the way for refined methodologies and continued efforts in promoting effective road safety measures, contributing to a safer transportation environment, and potentially reducing accident in Australia which will eventually lift the economy up.

# References

Australian Bureau of Statistics (2022). *Time series analysis: The Basics*. Retrieved from https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics.

Connelly, L.B., & Supangan, R. (2006). The economic costs of road traffic crashes: Australia, states, and territories. *Accident Analysis and Prevention, 38*(6), 1087-1093.

Damien, L. (1994). The cost of sleep-related accidents: A report for the National Commission on sleep disorders research (Special Report). *Sleep, 17*(1), 84-93.

Gorea, R. (2016). Financial impact of road traffic accidents on the society. *International Journal of Ethics, Trauma & Victimology, 2*(1), 6-9.

Hyndman, R.J., & Athanasopoulos, G. (2016). Forecasting: Principles and Practise. Retrieved from https://otexts.com/fpp2/simple-methods.html.

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfeld (Eds.), *International Handbook of Research in Statistics Education* (pp. 261–294). Springer.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Mannering, F.L., & Bhat, C.R. (2014). Analytic methods in accident research: methodological frontier and future directions. *Analytic Methods in Accident Research, 1*, 1-22.

Mgale, Y.J., Yan, Y.X., & Timothy, S. (2021). A Comparative Study of ARIMA and Holt-Winters Exponential Smoothing Models for Rice Price Forecasting in Tanzania. *Open Access Library Journal, 8*,1-9.

Jagnoor, J., Shankar, P., Lakshmi, P.V.M., Sameer, A., Belinda, G., & Rebecca, Q.I. (2015). The impact of road traffic injury in North India: A mixed-methods study protocol. *BMJ Open, 5*(8).

Zhang, X., Pang, Y., Cui, M., Stallones, L., & Xiang, H. (2015). Forecasting mortality of road traffic injuries in China using a seasonal autoregressive integrated moving average model. *Annals of Epidemiology,25*,101-106.

# Appendix

8.1. Young Adults Crashes

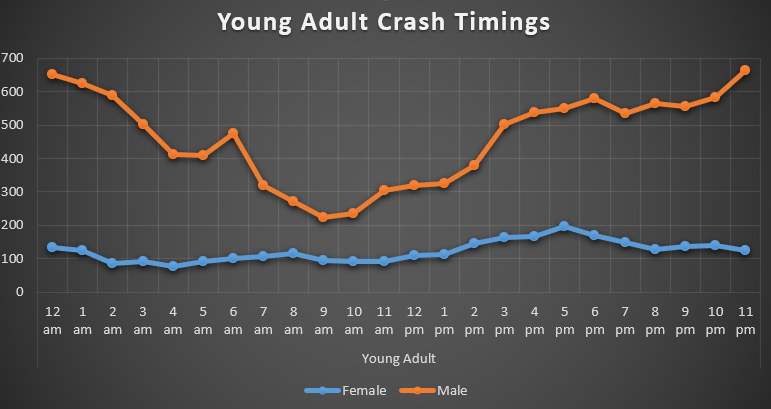


Figure : Crash timings of Young Adults

8.2. Age related distribution of crashes at the speed of 100.

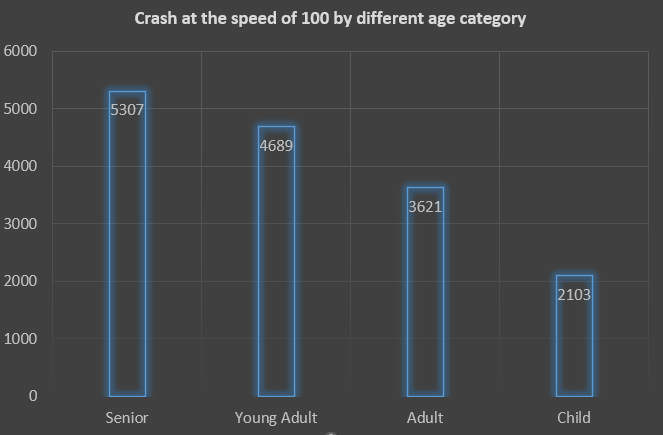


Figure : Involvement in crash by different Age-groups at the Speed of 100