Sales Analysis

Hamad S. Alyemni

SDAIA T5 Data Science and Artificial Intelligence Bootcamp

Applied Data Science in Python

Description of the data

The dataset is about sales of AAL, a clothing business in Australia. It contained the following six columns.

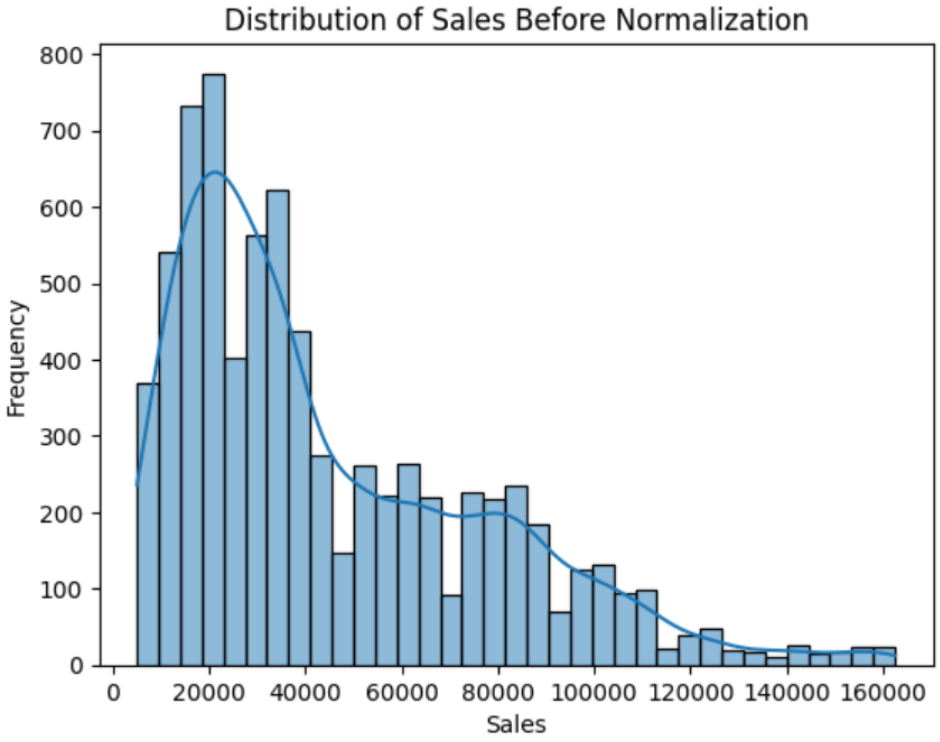
1. Date: stored in datetime data type in YYYY-MM-DD format.
2. Time: contains categorical times of the day. i.e., “Morning”, “Afternoon”, “Evening”.
3. State: a categorical abbreviation of the state name.
4. Group: contains the categorical names of client groups. i.e., “Men”, “Women”, “Seniors”, “Kids”,
5. Unit: contains the number of units sold
6. Sales: contains the amount of money generated in sales

The data consists of 7560 instances.

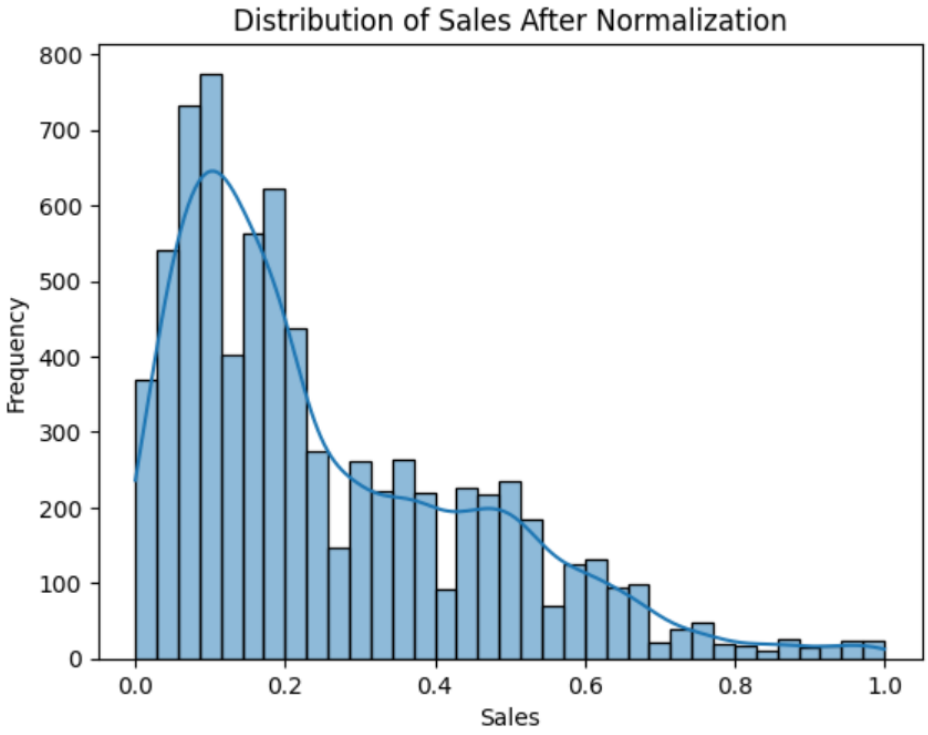
Data Wrangling

In terms of data wrangling and cleaning, the data was already clean. I checked for missing, duplicate, and incorrect values, and none were found.

I then checked the distribution of the data, and it was positively skewed. Meaning the data points were mostly clustered on the left half of the graph.



After normalization the distribution of the data points remained the same.



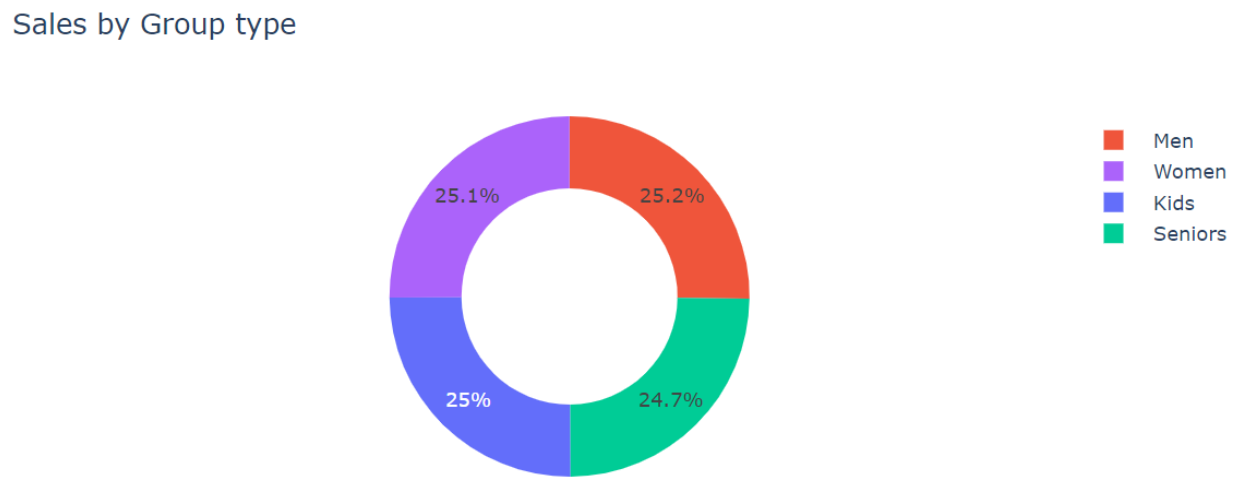
Data Analysis and Visualization

After the data wrangling, I started with the analysis of the data. I performed statistical analysis on the Unit and Sales columns to get a deeper understanding of the measures of central tendency.

	Unit	**Sales**
count	7560	7560
mean	18	45013
std	12.901403	32253
min	2	5000
25%	8	20000
50%	14	35000
75%	26	65000
max	65	162500

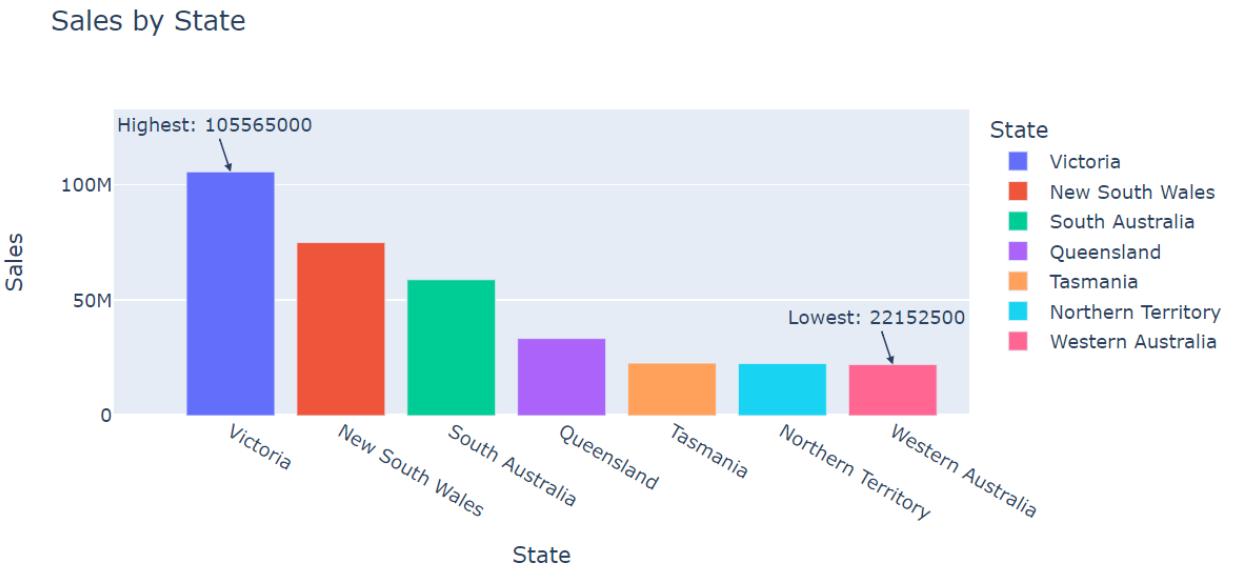
Afterwards, I visualized the data to answer some questions about it.

The first question is which group is generating the highest sales, and which group is generating the lowest sales?



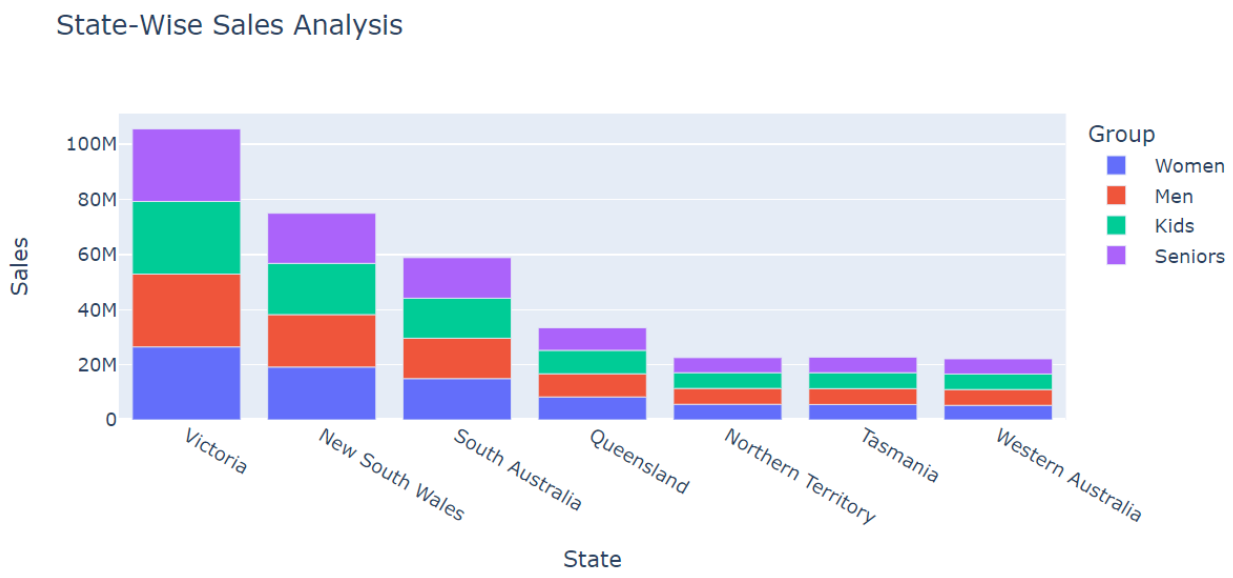
Men are the highest sale generating group, with $85,750,000 in sales while seniors are the lowest sale generating group, with $84,037,500 in sales. though the difference is small. coming to about $1,712,500

The second question is which state is generating the highest sales, and which state is generating the lowest sales?



Victoria is the highest sale generating state, with $105,565,000 in sales while Western Australia is the lowest sale generating state, with $22,152,500 in sales. The difference between the highest and lowest sale generating states is big. coming to about $83,412,500.

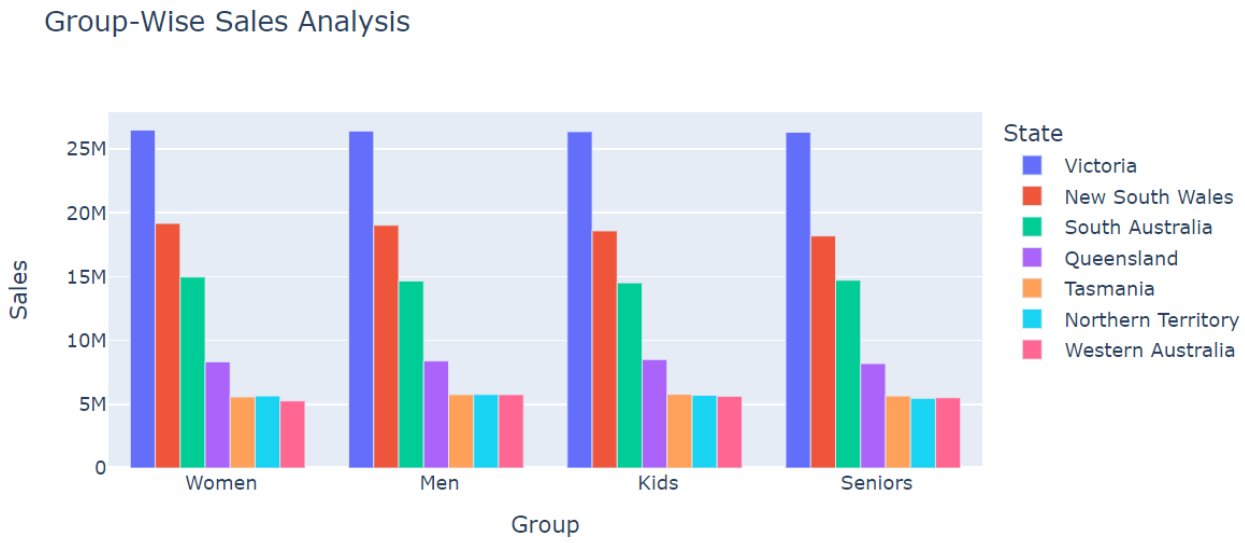
Then I visualized state-wise sales for different groups to get a deeper understanding of data.



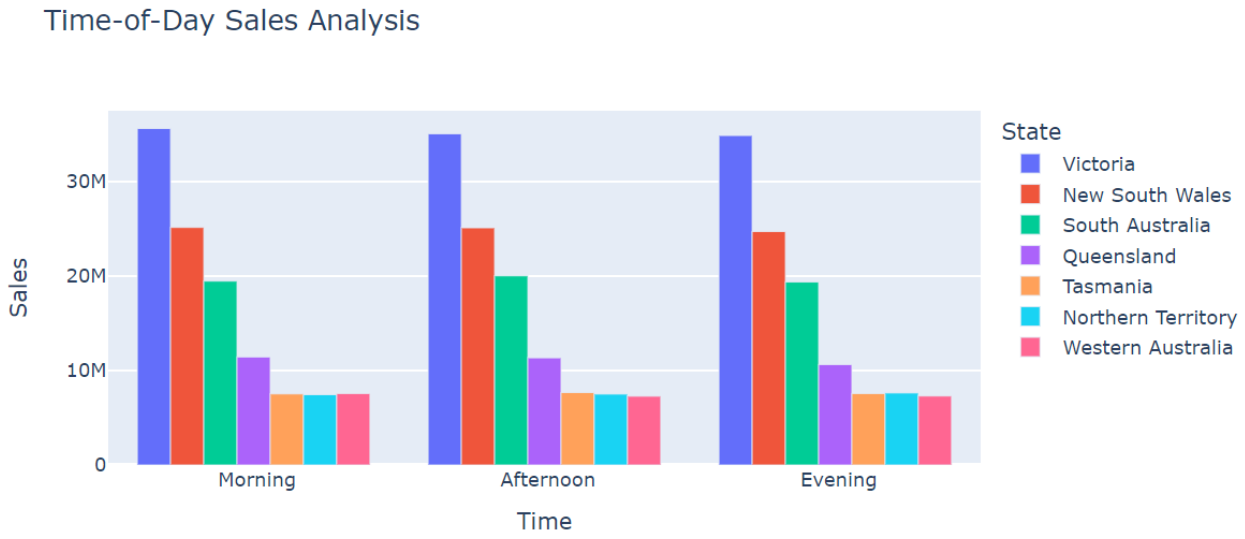
As we have seen before, Victoria is the highest sale generating state. and Western Australia is the lowest sale generating state.

Women generate the highest sale revenue and Seniors generate the lowest sale revenue in the state of Victoria while Men generate the highest sale revenue and Women generate the lowest sale revenue in the state of Western Australia.

Then I visualized group-wise sales for different states to get a deeper understanding of data.

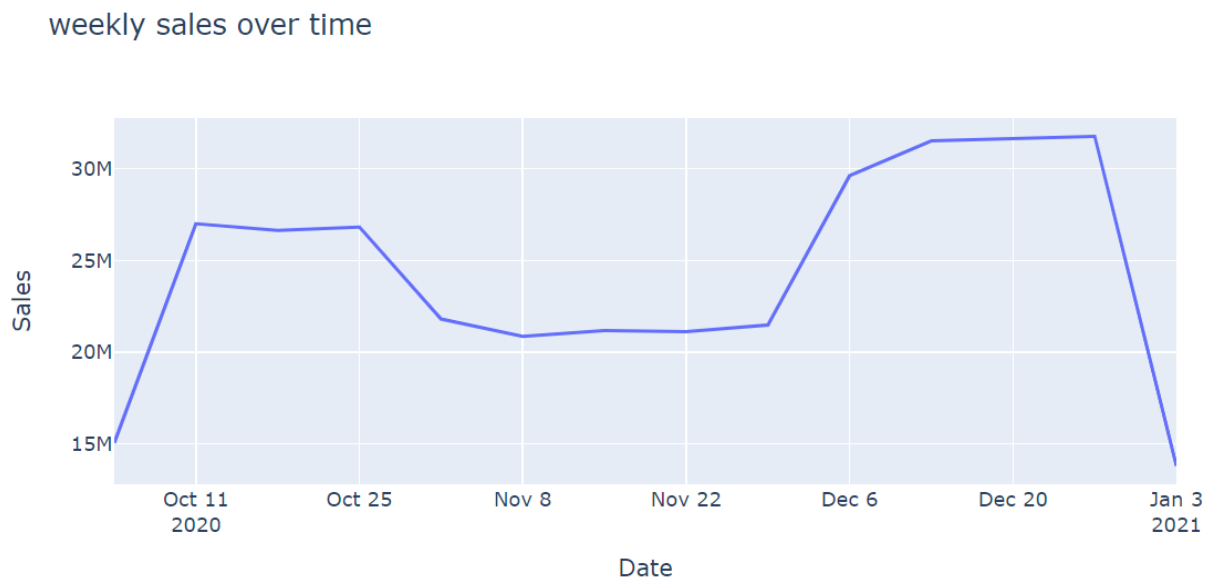


I also visualized time-of-day sales to understand which times of the day customers tend to shop at.

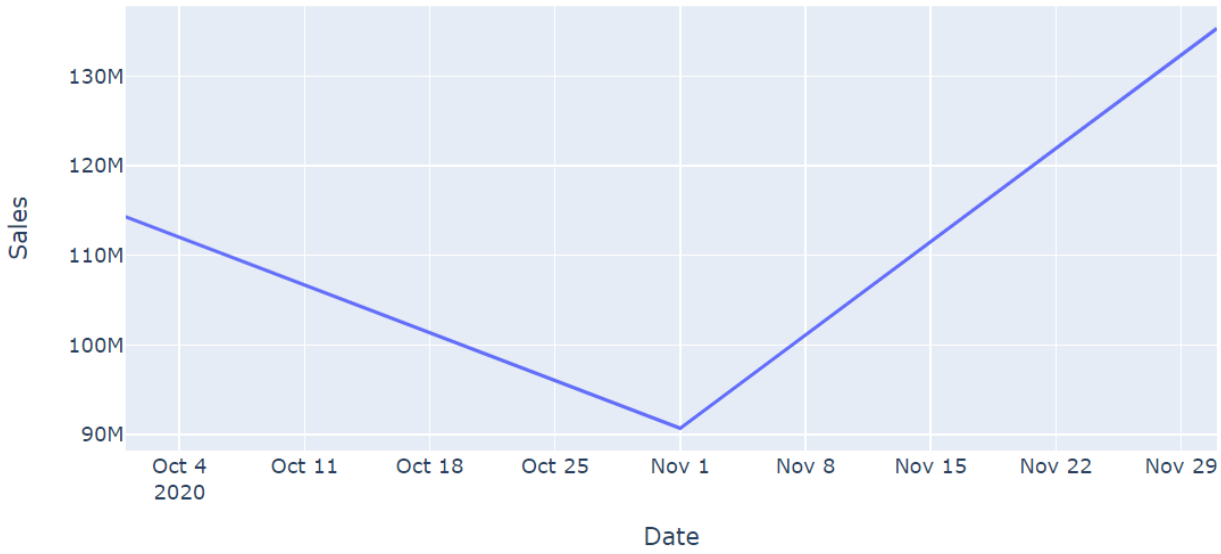


Revenue tends to be highest in the morning time in most states, and Revenue tends to be the lowest in the evening time.

Then I plotted the sales over time to highlight trends in sales.



We can see that the sales were at an all time low on October 4th 2020, and an all time high at December 27th 2020



We can see that the sales were the lowest on the month of November 2020, and the highest at December 2020