Time Series Modeling Summary (Non-Technical Report)

- In Canada, people tend to travel more during summer due to the weather, leading to higher cruise ticket sales. This seasonal pattern is clearly visible in the data, with noticeable spikes during summer each year, except in 2020, when travel was limited due to COVID-19. Overall, the data shows a consistent bell-shaped seasonal trend each year.
- During preprocessing, I noticed missing sales data for some days. The current code fills these missing days with zeros, which seems reasonable, as it likely means the business was closed on those days.
- The base model forecasts using the average sales by day of the year and returns a single value per time step. Also, it doesn't capture uncertainty, which is important for understanding forecast confidence.
- I improved the model by grouping the data using time related features like month, quarter, and day of the week. Finally, added the confidence intervals to reflect uncertainty.
- To predict sales, I applied a classical statistical model (SARIMA) and several machine learning models (XGBoost, Prophet, and DeepAR). I also used lag values and rolling averages as features in the models. Among all the models tested, XGBoost gave the best performance.
- With access to more external information, such as weather data, holidays, or promotions, the prediction accuracy could be further improved.