Question 1

1 Data Preparation

This section outlines the steps taken to clean and prepare the US Personal Consumption Expenditure (PCE) time series for forecasting. Given the importance of maintaining the temporal continuity of the series, particular attention was paid to handling missing values in a way that preserved underlying trends and structural components.

1.1 Identifying and Understanding Missing Data

Initial inspection of the dataset revealed 54 missing values in the PCE column out of 792 observations. A review of their locations showed no discernible temporal pattern or concentration, suggesting the data were *Missing Completely At Random* (MCAR). Since the dataset is a continuous monthly time series, removing rows with missing values would disrupt its temporal structure, negatively impacting modelling and decomposition. As a result, imputation was deemed necessary to preserve the integrity of the series.

1.2 Selecting the Best Imputation Method

Four imputation methods were considered, each selected for their relevance to time series data and ability to preserve trend dynamics:

1. **Linear Interpolation** — Estimates missing values using a straight line between known observations. Suitable for smooth transitions but may oversimplify dynamic series.
2. **Exponential Weighted Moving Average (EWMA)** — Averages prior values with exponentially decreasing weights. This method reflects short-term momentum but can dampen meaningful fluctuations.
3. **Kalman Smoothing (Structural Model: StructTS)** — Applies a state-space model that explicitly captures level and trend components. Particularly suited for economic data with consistent structural behaviour.
4. **Kalman Smoothing (ARIMA Model)** — Uses an automatically fitted ARIMA model to drive imputation.

1.3 Tuning Moving Average Imputation (k Selection)

The EWMA method required specifying a value for k, the window size that determines how many previous data points are averaged when imputing missing values. A smaller k captures local behaviour more responsively, while a larger k offers smoother, more stable results.

To select an appropriate value, autocorrelation (ACF) and partial autocorrelation (PACF) plots were examined using the `tsdisplay()` command.

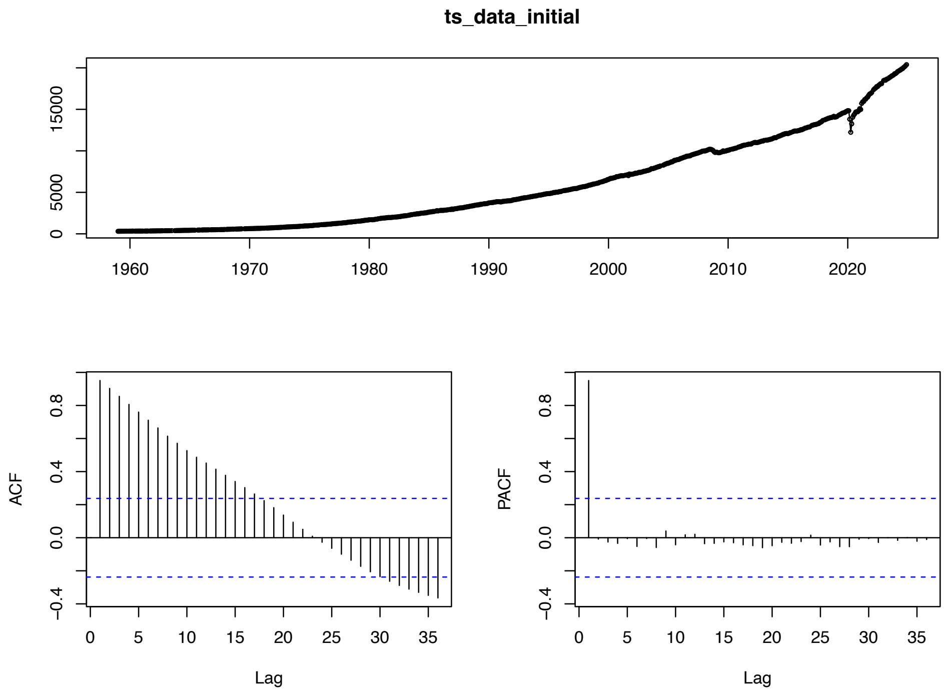


Figure 1: ACF and PACF plots used to determine optimal lag window size.

The ACF showed significant correlation up to lag 17, while the PACF dropped sharply after lag 1, suggesting that recent values carried strong predictive information. Based on this, a window of $k = 12$ — representing a full year of prior observations — was chosen to balance local responsiveness and trend preservation.

1.4 Method Comparison and Final Choice

All four imputation methods were applied to the original data, and the resulting series were visually compared with the original.

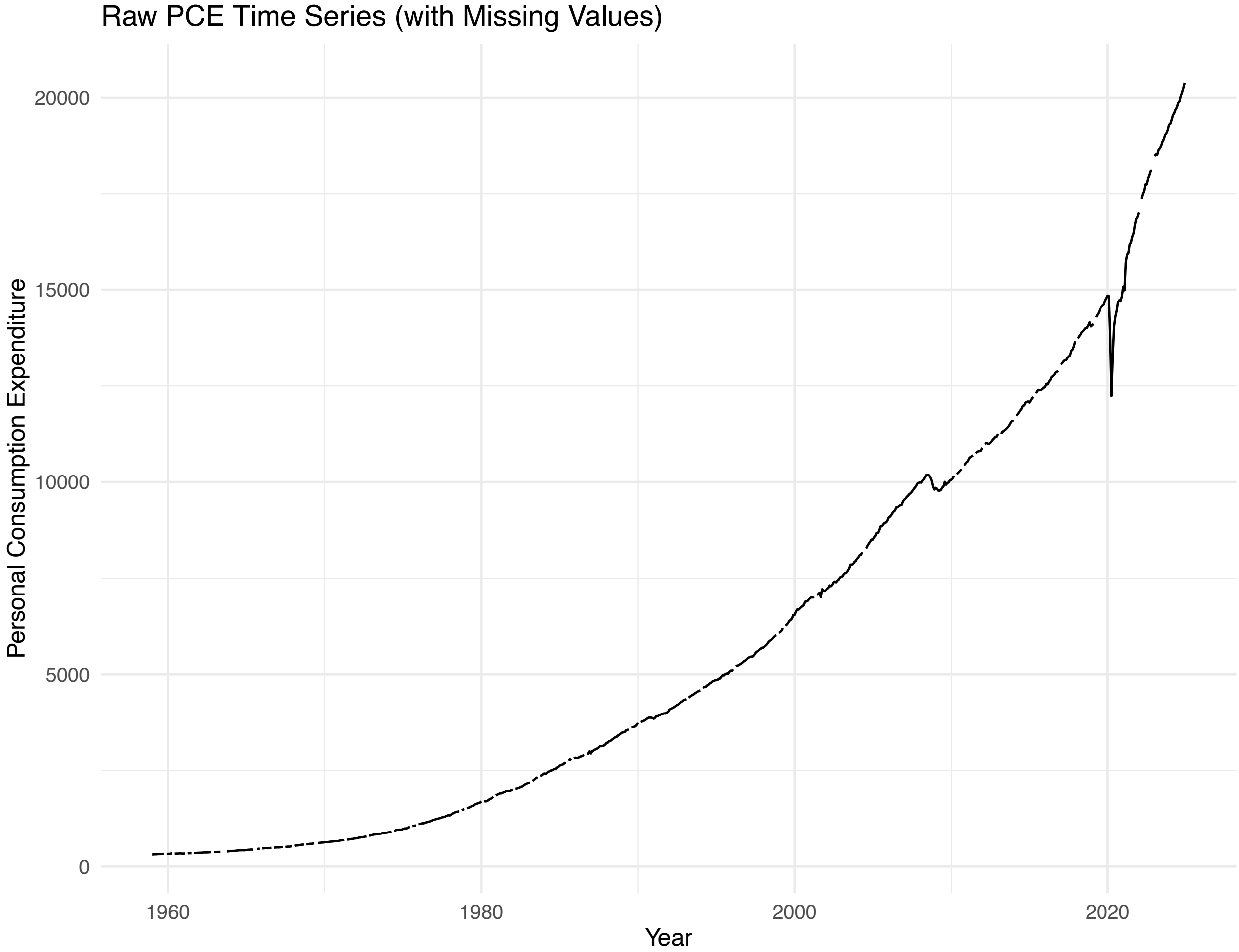
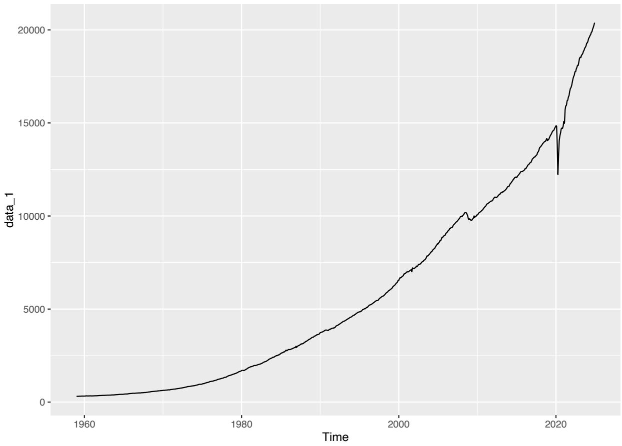
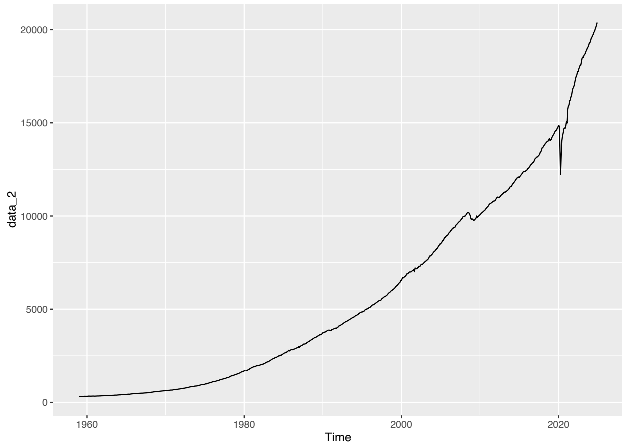


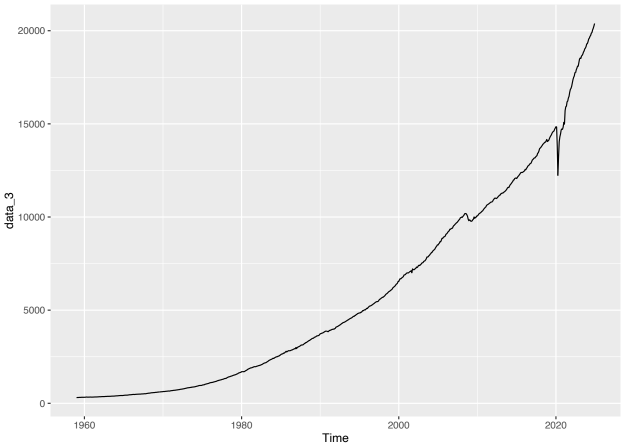
Figure 2: Original Time Series Plot (Before Imputation)



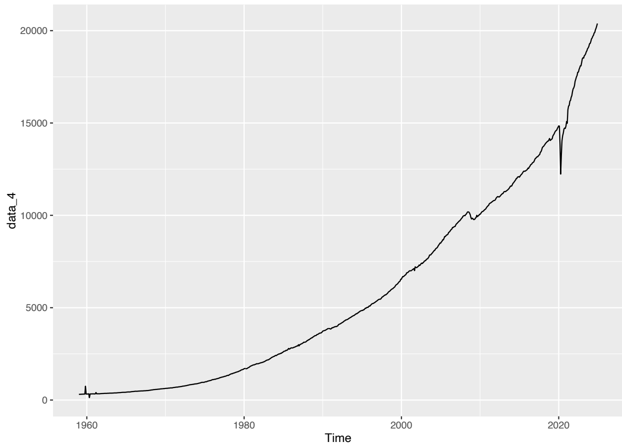
(a) Linear Interpolation



(b) EWMA (k = 12)



(c) Kalman Smoothing (StructTS)



(d) Kalman Smoothing (ARIMA)

Figure 3: Visual comparison of imputation methods.

The linear interpolation, EWMA, and `StructTS`-based Kalman smoothing methods all produced identical outputs. The ARIMA-based Kalman method, however, introduced anomalous spikes at the start of the series due to instability in early ARIMA fitting.

Since three of the methods yielded the same outcome, the final selection was made based on theoretical strength. The `StructTS` Kalman smoother was chosen as the most robust option due to its statistical foundation in state-space modelling and its ability to decompose level and trend components. This approach was best aligned with the long-term exponential growth seen in PCE data and avoided the instability introduced by the ARIMA-based method.

2 Exploratory Time Series Analysis

2.1 Trend Structure

Decomposition was first conducted using the `decompose()` function under an additive model assumption.

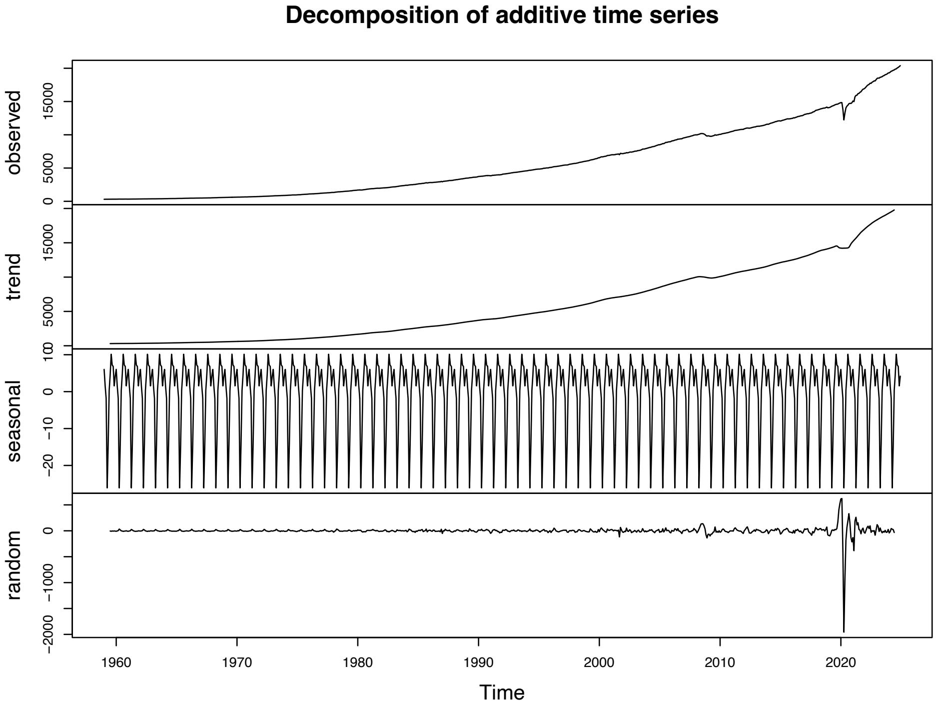


Figure 4: Additive decomposition of the PCE time series.

This revealed a smooth and continuous upward trend in PCE, consistent with long-term economic expansion. However, because the trend resembled exponential growth, a multiplicative decomposition was then applied. This approach is more appropriate when changes are proportional rather than additive.

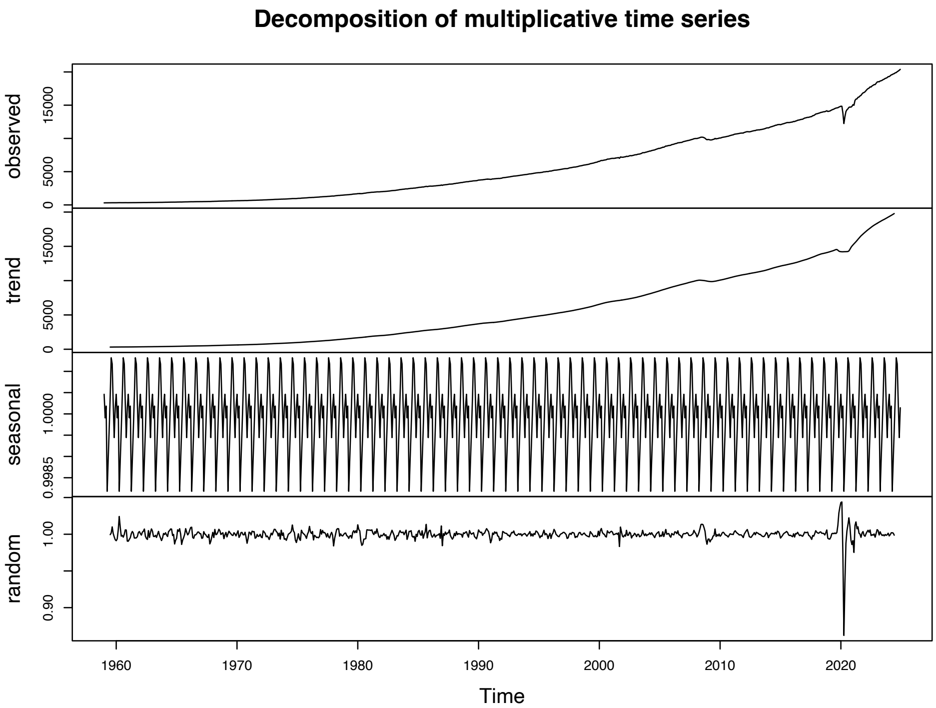


Figure 5: Multiplicative decomposition of the PCE time series.

The multiplicative decomposition confirmed the exponential trend and better reflected the nature of the series. The seasonal and random components showed minor fluctuations throughout the time, with the trend line clearly dominating. These findings suggest that the PCE series is primarily trend-driven, with growth accelerating over time.

2.2 Seasonality Assessment

To evaluate potential seasonality, several approaches were taken:

1. A seasonal plot of the original data showed little evidence of recurring monthly fluctuations. Lines representing different years were largely parallel, with no obvious monthly peaks or troughs. (See below)
2. Classical decomposition under both additive and multiplicative models revealed low-amplitude seasonal signals, suggesting a minor, consistent seasonal component. (See **Section 2.1**)
3. To isolate seasonality further, the data was first differenced to remove the trend, and a new seasonal plot was generated. This plot displayed no consistent monthly pattern, confirming that any seasonality in the original data was likely induced by the exponential trend. (See Below)

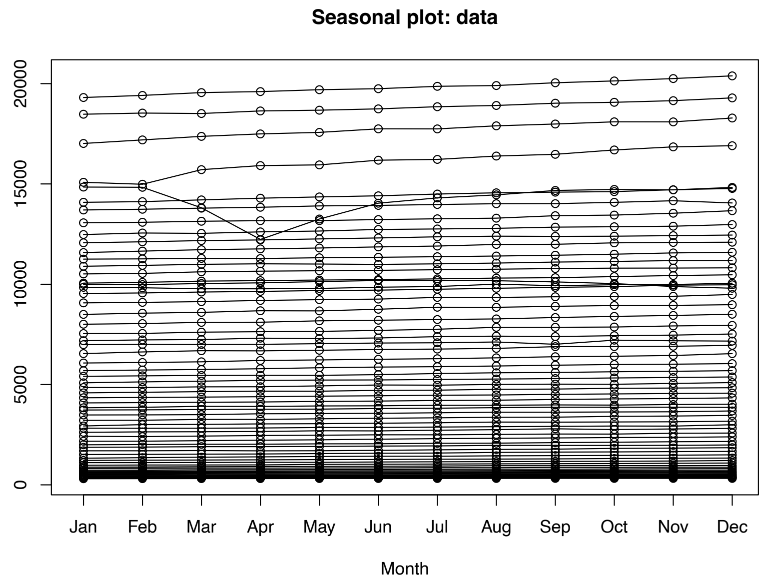


Figure 6: Initial Season plot.

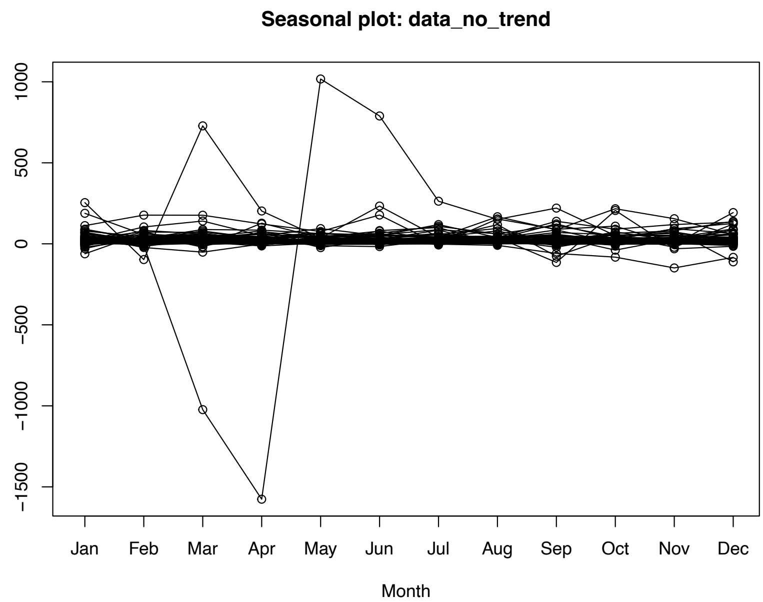


Figure 7: Season plot after first difference.

2.3 Conclusion on Data Characteristics

The analysis of the US Personal Consumption Expenditure (PCE) time series revealed several key structural features:

- **Exponential Trend:** The series is dominated by a strong exponential growth trend, reflecting long-term economic expansion. This was confirmed through decomposition and visual inspection of the data.
- **Minimal Seasonality:** While decomposition suggested a low-level seasonal signal, further investigation — including differencing and seasonal plots — confirmed that no meaningful, recurring monthly structure exists. Apparent seasonal components were likely artefacts of the dominant trend or external shocks.
- **Random Irregularities:** Deviations observed in the random component, particularly around 2020, are attributable to one-off shocks such as the COVID-19 pandemic. These anomalies do not represent structural features of the series and should be considered noise.
- **Non-Stationarity:** The presence of a trend and the absence of a constant mean and variance imply the series is non-stationary in its raw form. Differencing is necessary for models such as ARIMA to perform effectively.

Overall, the PCE time series can be characterised as non-stationary, non-seasonal, and trend-dominated. These properties inform the choice of forecasting models, guiding the selection toward approaches capable of capturing trend without overfitting non-existent seasonality.

3 Model Selection and Evaluation

The objective of this analysis was to evaluate and compare three forecasting approaches: a simple baseline method, an exponential smoothing model, and an ARIMA model. These were selected based on the specified criteria and their theoretical suitability for time series forecasting.

- **Naïve Forecasting Model:** Among various simple methods (e.g., mean forecast, seasonal naïve), the naïve model was chosen as it provides a straightforward benchmark by assuming future values equal the most recent observation. This makes it useful for assessing the added value of more complex models, particularly when dealing with non-seasonal data.
- **Exponential Smoothing Model (ETS):** The `ets()` function from the `forecast` package was used to automatically select the most appropriate exponential smoothing configuration. It considers all combinations of error, trend, and seasonality (additive or multiplicative), making it well-suited to the exponential trend structure found in the PCE data.
- **ARIMA Model:** The `auto.arima()` function was employed to determine the optimal ARIMA model using AICc minimisation. This automates differencing and parameter selection, yielding a statistically robust model. Given the non-stationary and trend-dominated nature of the series, ARIMA was a theoretically sound choice.

Together, these models represent a spectrum from simple to advanced, allowing for a comprehensive evaluation of forecast accuracy and robustness.

3.1 Pre-Processing and Train-Test Split

To evaluate forecast performance, the data was split into training and test sets. The training set included all data up to December 2023, while the test set comprised the 12 months of 2024. This split aligns with the business objective of forecasting the year ahead to support strategic planning. Assessing model accuracy on the most recent period provides a strong indication of its predictive reliability for future values.

3.2 Evaluation Metrics

To assess model performance, several standard forecasting accuracy metrics were used:

- **ME (Mean Error):** Measures the average forecast bias. Positive values indicate under prediction, negative values indicate overprediction.

- **RMSE (Root Mean Squared Error):** Penalises large forecast errors more heavily by squaring them before averaging, making it sensitive to outliers.
- **MAE (Mean Absolute Error):** Calculates the average absolute difference between forecasts and actual values.
- **MAPE (Mean Absolute Percentage Error):** Expresses forecast error as a percentage of actual values, making it useful for interpretability across scales.
- **MASE (Mean Absolute Scaled Error):** Compares forecast performance against a naïve benchmark. Values less than 1 indicate better-than-naïve performance.
- **ACF1 (Lag-1 Autocorrelation of Residuals):** Assesses whether residuals are autocorrelated. Values close to 0 suggest residuals resemble white noise which implies good model fit.
- **Theil's U:** A relative accuracy measure compared to a naïve model. Values less than 1 indicate improved forecasting performance.

3.3 Naïve Model

The naïve model assumes that the most recent observation is the best predictor of future values. Therefore, the naïve forecast always produces a flat line across the prediction period stemming from the last actual vlaue.

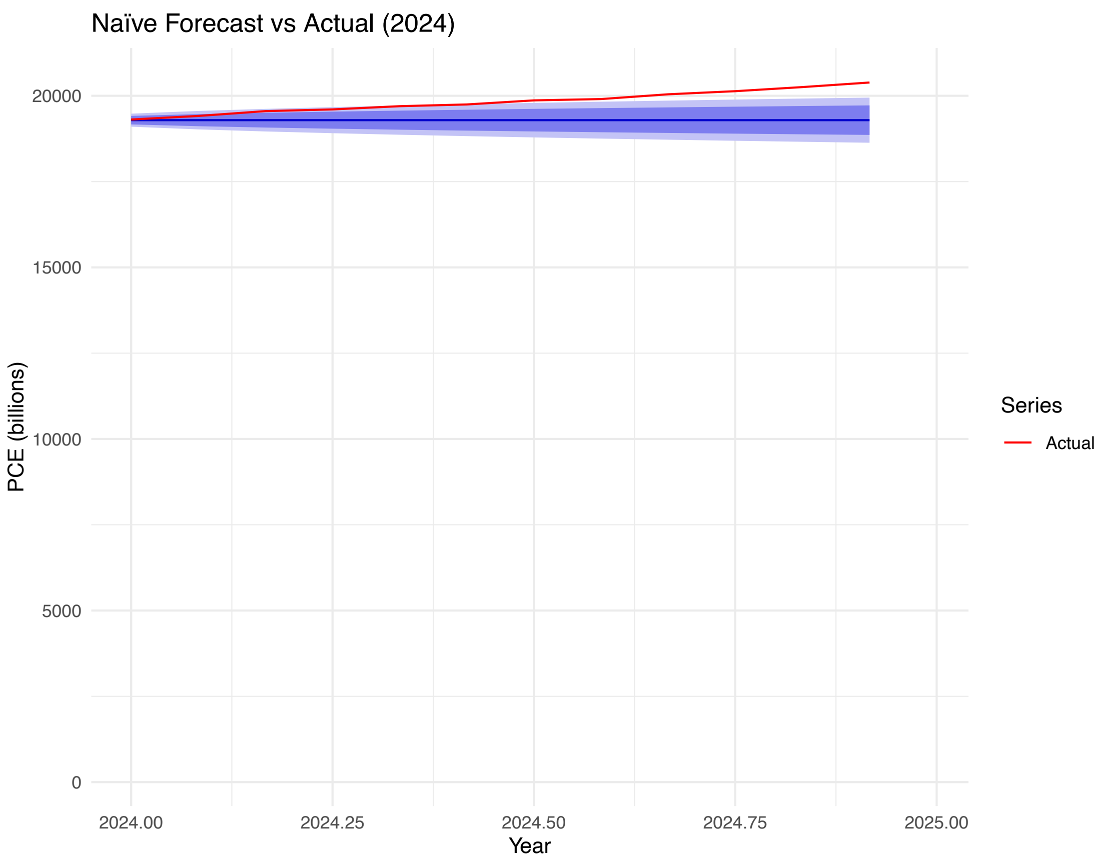


Figure 8: Naive Model Forecast vs Actual.

The actual PCE values in 2024 showed continued growth, leading the naïve model to consistently underpredict. This resulted in a high mean error (ME = 536.2) and a MAPE of 2.68%, indicating poor forecast accuracy. The residuals displayed strong autocorrelation ($\text{ACF}_1 = 0.72$), suggesting unmodelled structure in the data.

While useful as a reference, the naïve model is unable to capture the exponential trend and was outperformed across all accuracy metrics by the ETS and ARIMA models.

3.4 Exponential Smoothing (ETS) Model

The selected model was ETS(M,A,N), featuring multiplicative errors, an additive trend, and no seasonal component — a structure well suited to non-seasonal data with exponential growth. This aligns closely with the observed characteristics in the PCE series.

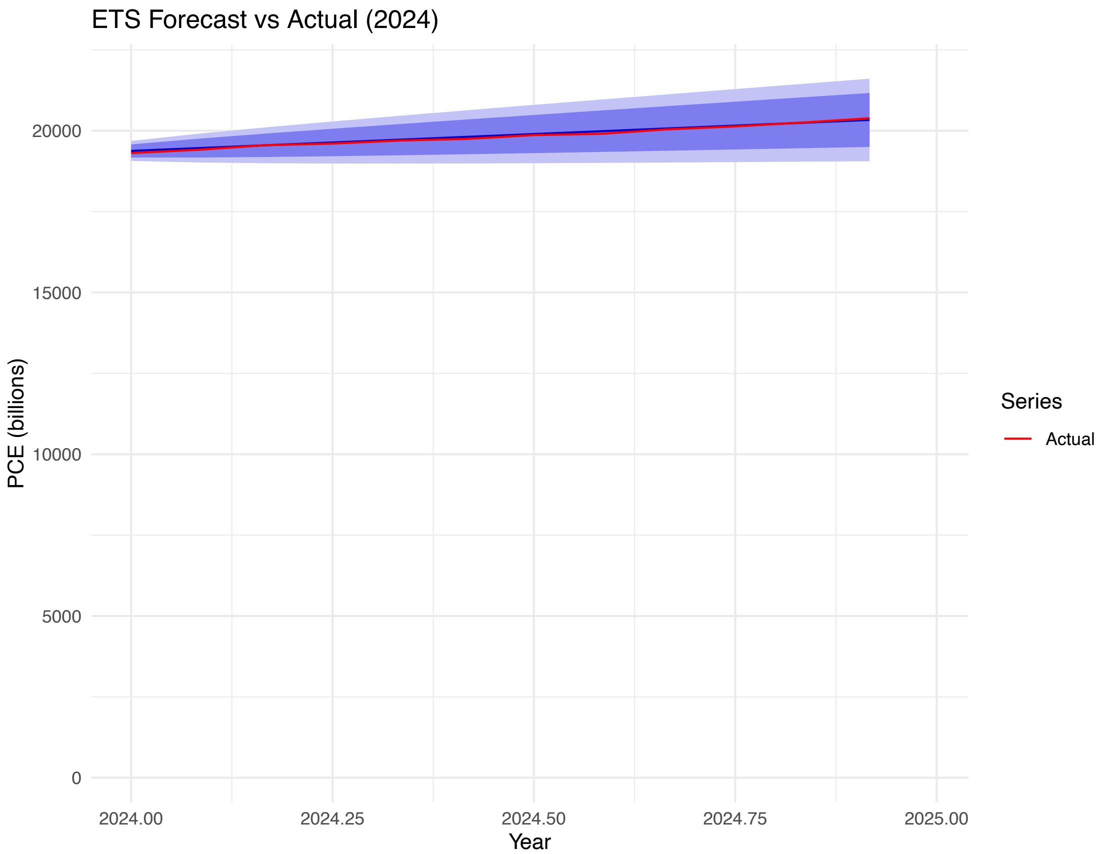


Figure 9: Exponential Smoothing Model Forecast vs Actual.

The forecast closely followed actual values throughout 2024, capturing the trend and remaining within the prediction intervals. With a MAPE of 0.20% and MASE of 0.13, ETS substantially outperformed the naïve model, confirming its suitability for short-term economic forecasting.

3.5 ARIMA Model

The ARIMA model was fitted using `auto.arima()`, which selected ARIMA(2,2,1)(0,0,1)[12] based on AICc. This specification includes two autoregressive terms (AR), one moving average term (MA), second-order differencing ($d = 2$), and a seasonal MA term with a 12-month period.

Second-order differencing was necessary to stabilise the exponential trend, and the seasonal MA term captured minor autocorrelated structure. Despite no strong seasonal pattern, this hybrid configuration enabled the model to account for both long-term trend and short-term irregularities.

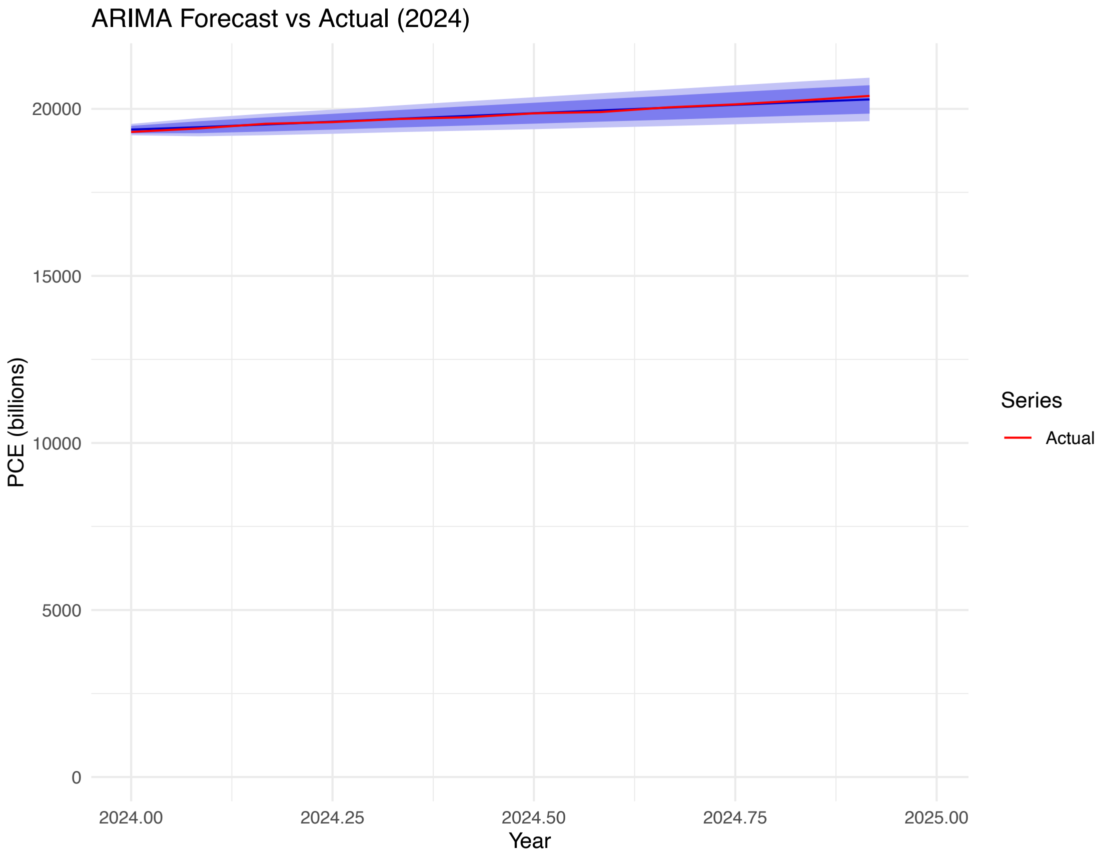


Figure 10: ARIMA Model Forecast vs Actual.

The forecast tracked actual values closely throughout 2024, with tight prediction intervals and minimal deviation.

3.6 Forecast Accuracy Comparison & Model Recommendation

Table 1: Test Set Performance Metrics by Forecasting Method

Metric	**Naïve**	**ETS (M,A,N)**	**ARIMA(2,2,1)(0,0,1)[12]**
ME	536.20	-29.26	-1.89
RMSE	624.76	45.92	45.49
MAE	536.20	39.63	34.79
MAPE (%)	2.68	0.20	0.18
MASE	1.75	0.13	0.11
ACF1	0.72	0.22	0.31
Theil's U	6.19	0.42	0.40

The naïve model, while serving as a baseline, failed to reflect the strong trend in PCE, leading to large errors. ETS significantly improved forecast accuracy however, the ARIMA model consistently produced the most accurate forecasts, achieving the lowest errors across most metrics.

Conclusion: Based on both theoretical appropriateness and empirical performance, the ARIMA(2,2,1)(0,0,1)[12] model is the preferred choice for forecasting Personal Consumption Expenditures in 2025.

4 Final Forecast and Predictions

The goal was to determine the most accurate model for forecasting monthly U.S. Personal Consumption Expenditure (PCE) values for 2025. Table 2 presents the monthly forecasts generated by each of the three evaluated models. Based on prior analysis, the ARIMA model produced the most reliable predictions and is recommended for guiding business strategy.

4.1 Twelve-Month Forecast for 2025

Table 2: Forecasted PCE for 2025 by Model (in billions)

Month	Naïve	ETS	ARIMA
Jan	20387.2	20475.86	20480.44
Feb	20387.2	20564.66	20550.40
Mar	20387.2	20653.47	20628.56
Apr	20387.2	20742.28	20720.65
May	20387.2	20831.09	20807.96
Jun	20387.2	20919.90	20896.60
Jul	20387.2	21008.70	20978.86
Aug	20387.2	21097.51	21069.20
Sep	20387.2	21186.32	21150.31
Oct	20387.2	21275.13	21236.16
Nov	20387.2	21363.93	21319.16
Dec	20387.2	21452.74	21400.28

4.2 Visual Comparison of Forecasts

Figure 11 presents a visual comparison of the overlaid model predictions.

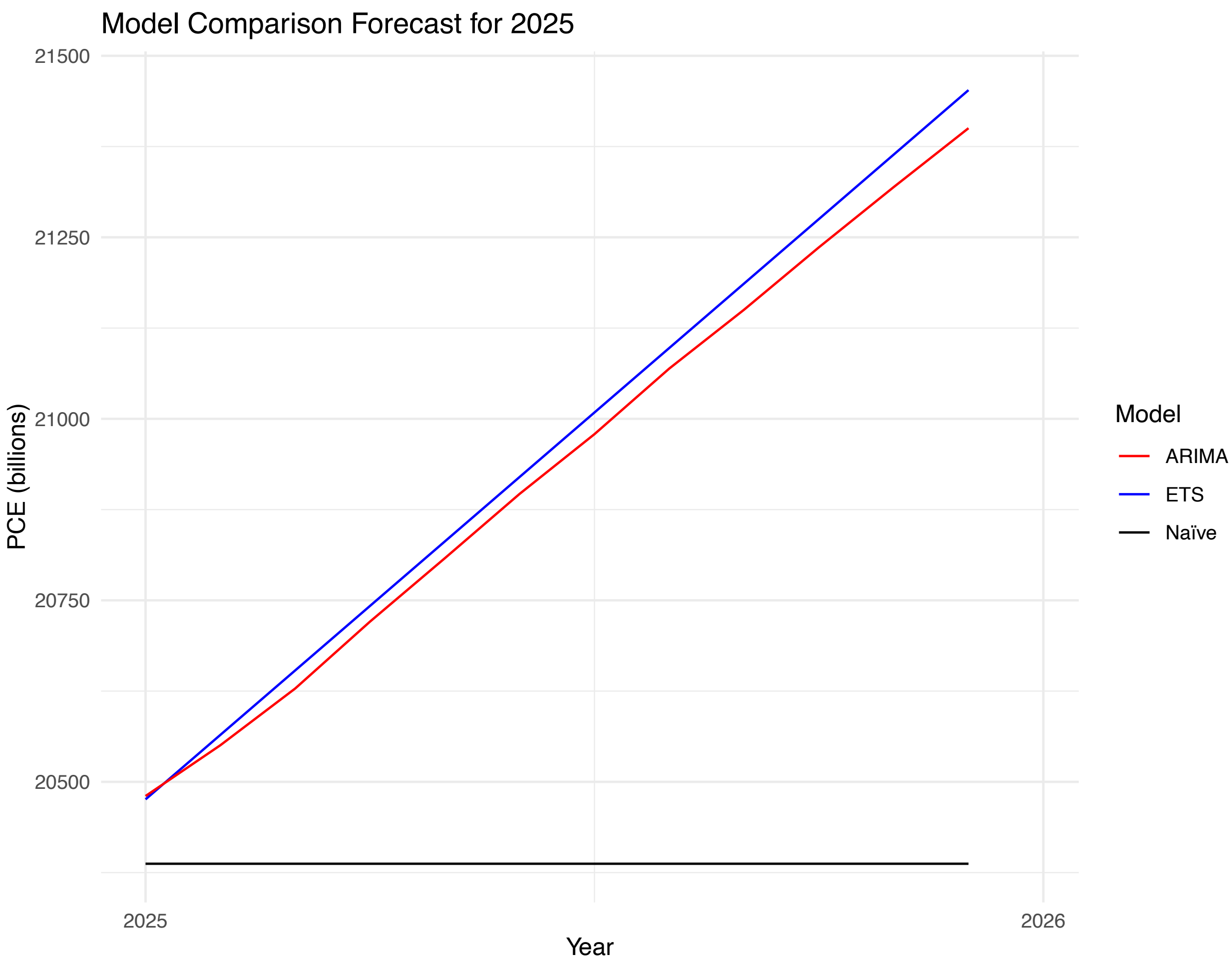


Figure 11: Comparison of Predictions from each Model.

Question 2

1 Data Preparation

To ensure that the analysis delivered clear and actionable insights, the hotel reviews dataset underwent a rigorous cleaning process. Each record included a satisfaction score (1 to 5) and a free text review describing the customer's experience. A series of filtering steps was applied to ensure that only relevant and reliable reviews were included.

1.1 Removal of Neutral Reviews

Reviews with a satisfaction score of 3 were excluded to improve the contrast between satisfied and dissatisfied customers. A score of 3 typically reflects neutral sentiment and could introduce ambiguity. By focusing on ratings of 1–2 (negative) and 4–5 (positive), the analysis more clearly identified the factors influencing customer satisfaction or dissatisfaction.

1.2 Filtering for English-Language Reviews

The original dataset contained reviews in multiple languages, so a language detection algorithm was used to retain only English entries. This ensured consistency, as topic modelling techniques rely on recognising language patterns that could be distorted by multilingual content.

1.3 Sampling for Efficiency and Reproducibility

Due to the computational demands of topic modelling, a random sample of 2,000 English-language reviews was selected. A fixed random seed was applied to maintain reproducibility, allowing the results to be replicated exactly for analytical integrity and future validation.

2 Text Cleaning and Preparation

Following the initial review sampling, two separate datasets were created corresponding to positive (scores of 4–5) and negative (scores of 1–2) reviews. Each dataset was used to build its own corpus and Document-Term Matrix (DTM). This separation enabled me to independently model the themes present in positive and negative feedback.

Initial attempts to use the `DocumentTermMatrix()` function's built-in cleaning parameters (e.g., `tolower`, `removePunctuation`, `stopwords`) were insufficient, as they did not reliably handle punctuation within words or special characters. Therefore, I implemented a more robust, manual cleaning pipeline to improve reliability and transparency.

All review text was first converted to UTF-8 encoding to ensure compatibility and remove problematic characters that can arise from mixed-language inputs or web-sourced content. I then removed non-standard punctuation, hyphen variants, and non-ASCII symbols using regular expressions.

After this initial step, I created two separate corpora—one for positive reviews and one for negative—and applied further cleaning using the `tm_map()` function from the `tm` package. This included lowercasing, punctuation and number removal and white-space stripping.

Stopword filtering using a standard English list and word-level lemmatisation were also applied to simplify the review text. **Stopwords** are common words such as *"the"*, *"is"*, and *"and"* that appear frequently but carry little meaning. Removing these helps focus the analysis on more meaningful content. **Lemmatisation** is the process of converting words to their base form (e.g., *running* $\rightarrow$ *run*) to group similar terms and reduce vocabulary size. This step ensures that semantically related words are treated consistently, without the distortion that can occur with more aggressive techniques like stemming.

DTMs were then constructed from the cleaned corpora, and any documents left empty after preprocessing were removed.

Finally, to validate the cleaning process, I manually inspected the resulting term lists using regular expression-based searches. I specifically checked for residual punctuation, numbers, and capitalised words, confirming that the vocabulary had been successfully standardised and noise removed.

3 Exploratory Word Cloud Visualisations

To gain an initial understanding of the terms most frequently discussed in the hotel reviews, separate word clouds were generated for the positive and negative datasets. These visualisations display the most common terms based on raw frequency within each sentiment group, with word size corresponding to the number of times each term appears.

In both word clouds, core hotel-related terms such as *hotel*, *room*, and *stay* dominate, which is expected given the review context. However, more interesting insights emerge from the surrounding terms.

3.1 Negative Word Cloud

The word cloud for negative reviews highlights terms such as *shower*, *bed*, *bathroom*, *clean*, *reception*, *price*, and *small*. These indicate key dissatisfaction areas: poor maintenance and cleanliness, especially in bathrooms; discomfort or limited space in rooms; underwhelming front desk experiences; and perceptions of poor value for money. Addressing these issues could significantly improve customer satisfaction.

Figure 12: Negative review Word Cloud.

3.2 Positive Word Cloud

In contrast, the positive word cloud highlights terms such as *friendly*, *tube*, *station*, and *comfortable*. These suggest that guests value welcoming staff, convenient access to public transport, and a comfortable stay. Enhancing these strengths could reinforce positive guest experiences and loyalty.

Figure 13: Positive review Word Cloud

These word clouds serve as an effective exploratory tool, providing a high-level overview of review content and helping to validate the decision to model the two sentiment groups separately.

4 Topic Modelling using LDA

To identify the key themes discussed in positive hotel reviews, I applied Latent Dirichlet Allocation (LDA), a probabilistic topic modelling technique that uncovers latent semantic structures based on word co-occurrence patterns. This method provides an interpretable summary of the underlying topics in the corpus without relying on predefined labels.

4.1 Topic Tuning Process

Before fitting the LDA models, I used the `ldatuning` package to determine the optimal number of topics (k) for each review corpus. This involved fitting LDA models across a range of topic numbers from 5 to 20 using Gibbs sampling, and evaluating model quality using three diagnostic metrics:

- **Griffiths2004**: A metric based on log-likelihood, which should be maximised to indicate better model fit.
- **CaoJuan2009 and Arun2010**: Two divergence-based metrics that should be minimised to ensure distinct and well-separated topics.

The values of these metrics were then visualised using the `FindTopicsNumber_plot()` function to identify the topic number that best balanced model performance and interpretability. This tuning process was carried out separately for the positive and negative review corpora, allowing the LDA models to be tailored to the specific characteristics of each dataset.

4.2 Optimal Number of Topics – Positive Reviews

The positive review corpus was modelled using 14 topics, based on the results of the tuning process.

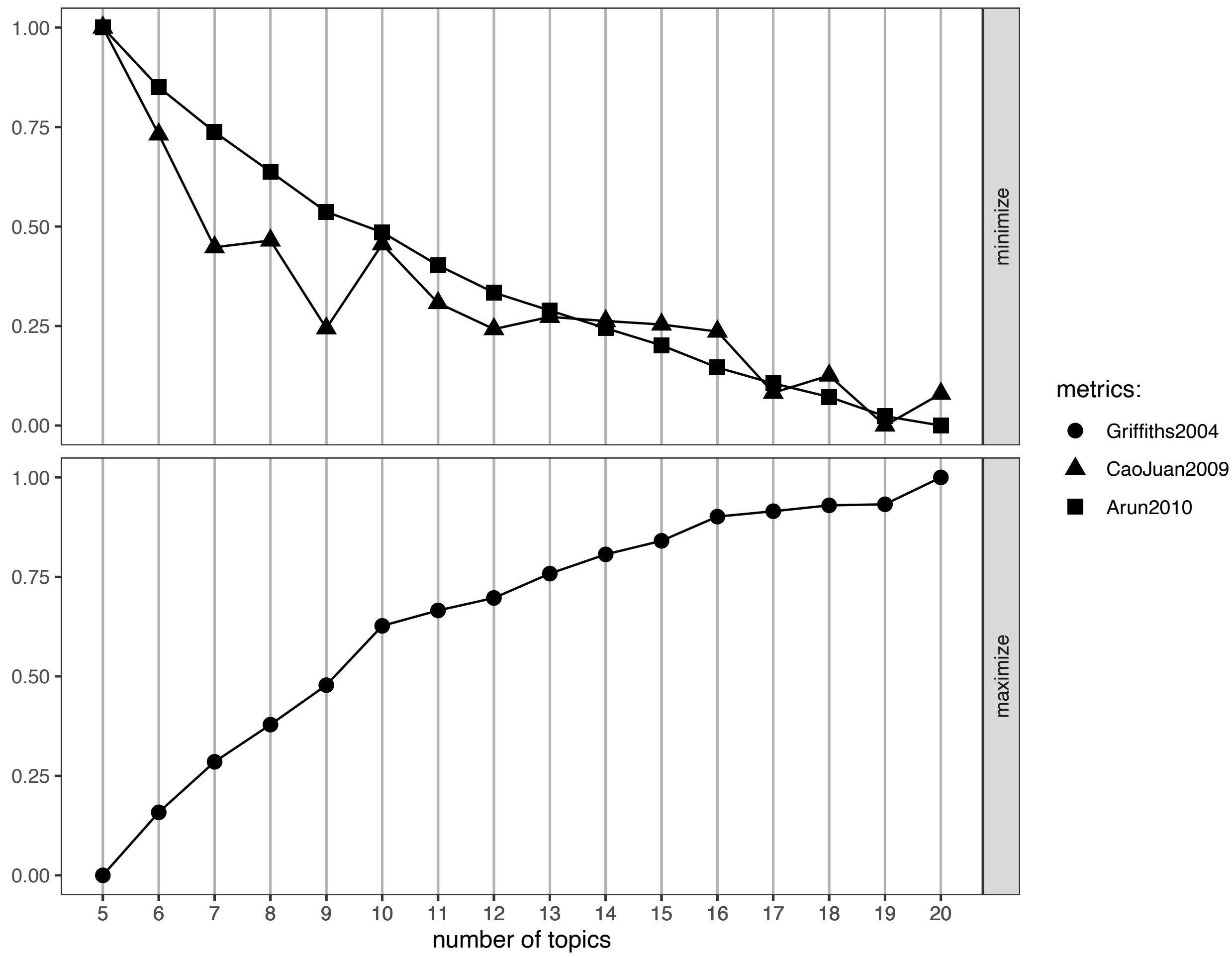


Figure 14: Tuning Process for Positive Reviews Topic Modelling

This choice ensured that the model captured sufficient thematic diversity without overfitting or producing overly granular, redundant topics. Each of the 14 topics was manually interpreted by reviewing its top associated terms and grouping them into meaningful themes.

Key themes that emerged from the positive reviews included:

- Staff friendliness and cleanliness
- Breakfast satisfaction and dining quality
- Proximity to public transport and major attractions
- Room comfort, features, and sleep quality
- Overall experience with check-in and front desk service

These topics reflect the multifaceted nature of guest satisfaction and highlight the specific aspects of the hotel experience that were most frequently praised.

4.3 Topic Prevalence in Positive Reviews

To assess which themes were most commonly discussed, I calculated the average proportion of each topic across all reviews. The resulting bar chart (Figure 15) shows how frequently each topic appeared.

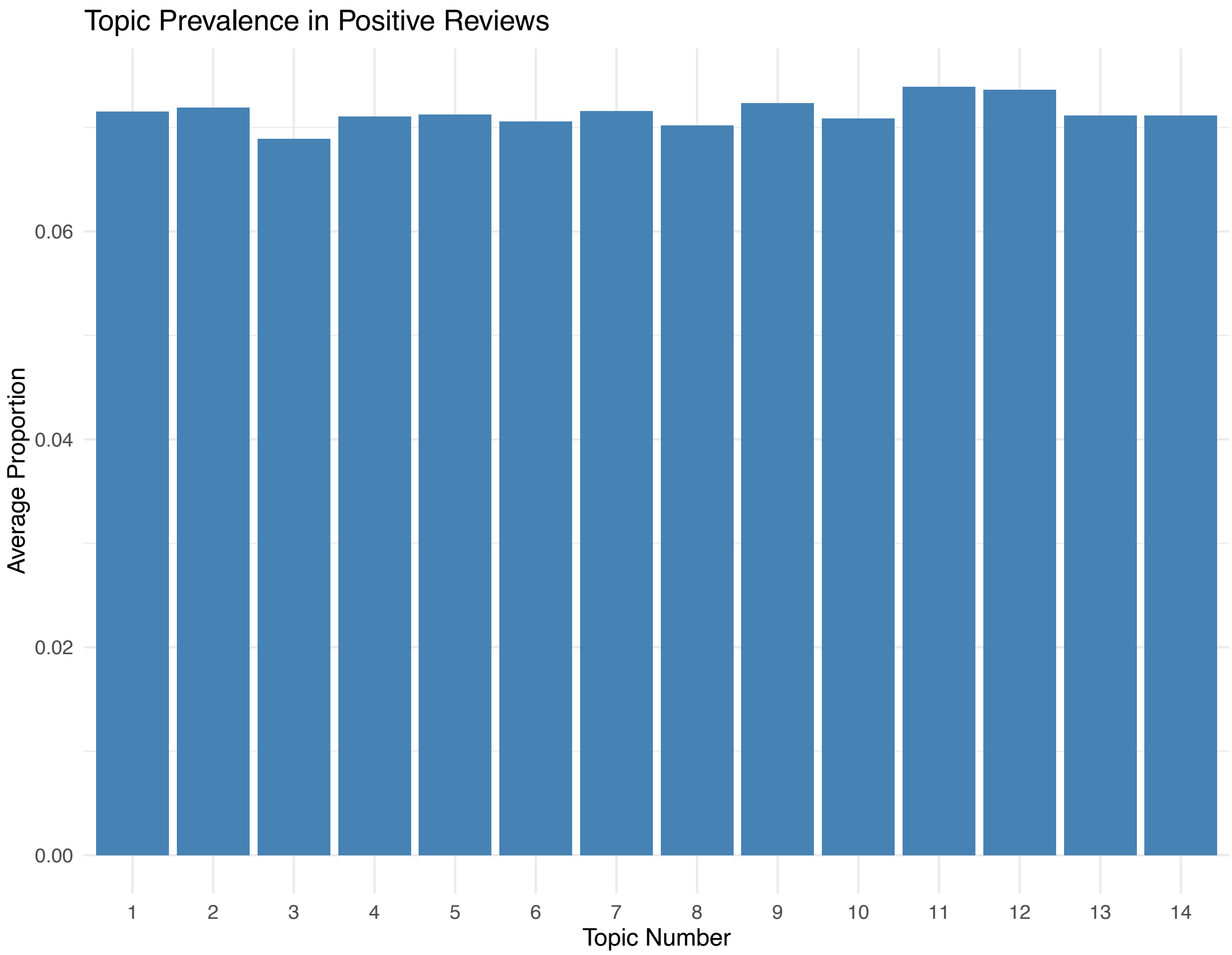


Figure 15: Topic Prevalence Amongst Positive Reviews

The three most prominent topics were:

- **Public Transport Access (Topic 12)** – reflecting location convenience and proximity to stations or underground lines, with terms like *station*, *tube*, *walk*, and *close*
- **Friendly Staff & Cleanliness (Topic 9)** – highlighting satisfaction with helpful, professional staff and clean, comfortable rooms, using words such as *staff*, *clean*, *friendly*, and *welcome*
- **General Hotel Experience (Topic 1)** – reflecting overall impressions of the hotel's quality, amenities, and suitability for business or leisure, with common terms like *hotel*, *nice*, *bar*, and *modern*

Although Topic 11 (*Experience & Atmosphere*) had a relatively high prevalence, it was excluded from this summary due to its lack of specificity regarding tangible aspects of the hotel experience. Its terms reflected vague sentiment rather than actionable themes.

Overall, the results suggest that accessibility, cleanliness and staff service, and overall hotel quality are central to positive guest experiences.

4.4 Optimal Number of Topics – Negative Reviews

The negative review corpus was modelled using 15 topics, selected based on the output of the tuning process.

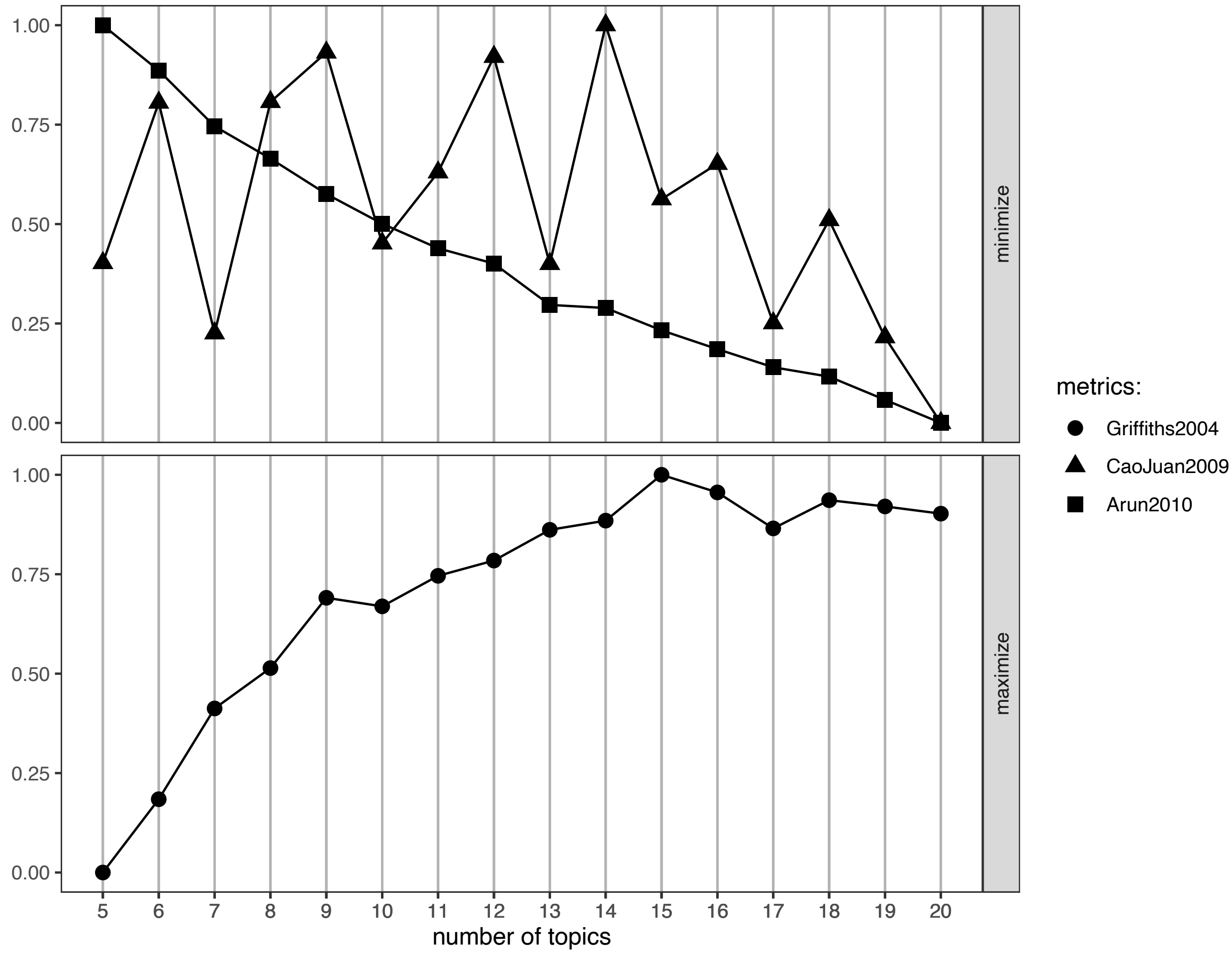


Figure 16: Tuning Process for Negative Reviews Topic Modelling

The `Griffiths2004` metric increased consistently and plateaued around 15 topics, while both `CaoJuan2009` and `Arun2010` metrics showed a steady decline, indicating improved topic separation and model quality.

4.5 Topic Prevalence in Negative Reviews

To explore which issues were most commonly raised in negative reviews, the average topic probability across all documents was calculated as before and the results displayed in the following bar chart.

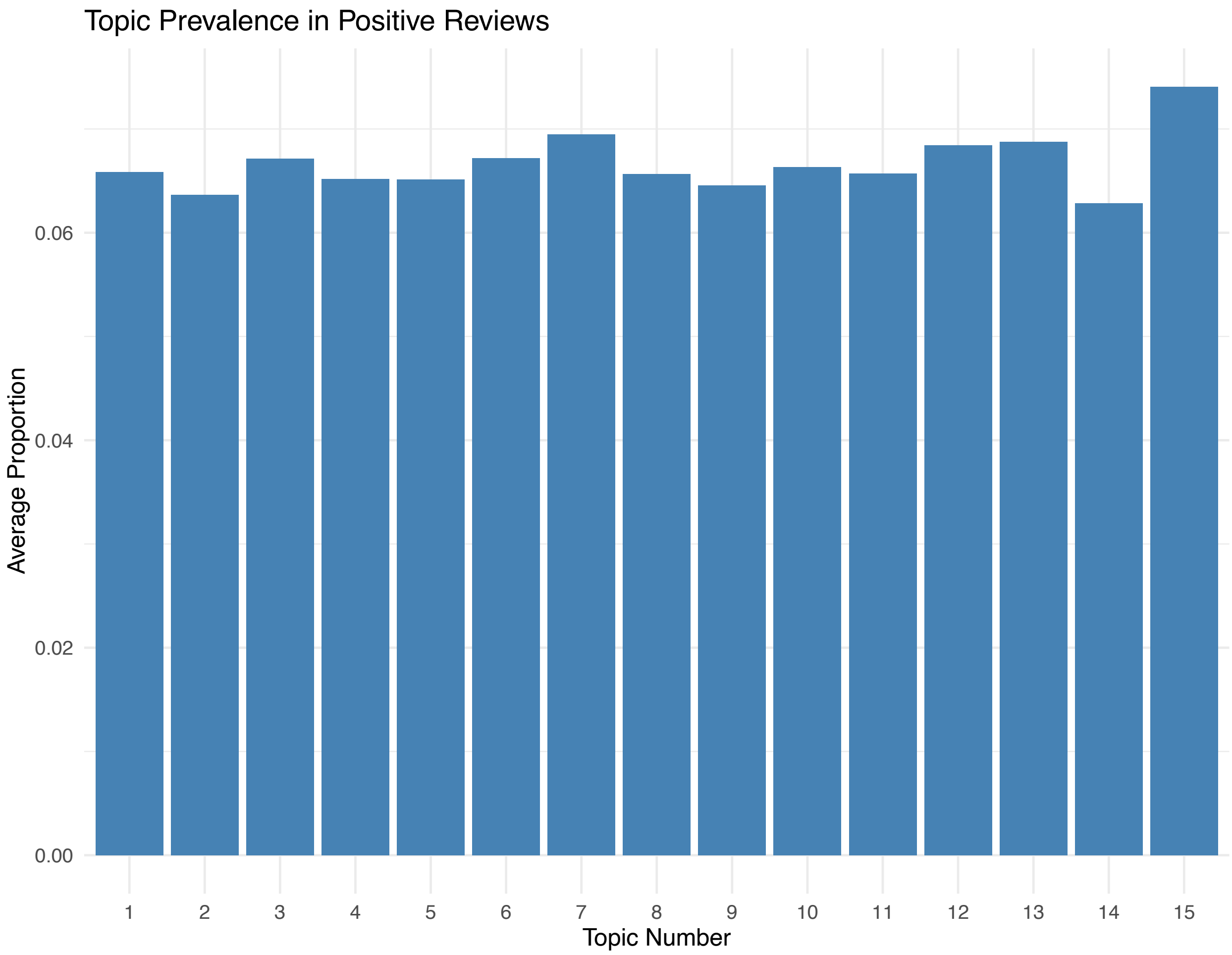


Figure 17: Topic Prevalence Amongst Negative Reviews

While topic proportions were generally well distributed, **Topic 15** emerged as the most prevalent. This topic included a mixture of faint praise for elements like breakfast, staff, or location—often contrasted against broader complaints—suggesting that customers occasionally mentioned small positives even within otherwise negative reviews.

Other notably prominent themes included:

- **Topic 9**: Issues with temperature, water, or reception—such as cold showers or poor front desk communication
- **Topic 1**: Unmet expectations regarding quality or value relative to the hotel's star rating

The remaining topics covered a wide range of dissatisfaction points, including room cleanliness, broken facilities, noise, unresponsive service, and check-in problems. The relatively even spread indicates that guest dissatisfaction is rarely driven by a single issue but rather by a combination of factors—many of which relate to service quality and room condition.

5 Business Recommendations

1. **Prioritise maintenance and repair processes.**
   Many complaints focused on broken fixtures, showers, heating/air systems, and general wear-and-tear. A routine, well-documented inspection and repair cycle could reduce these complaints significantly.
2. **Enhance staff responsiveness and training.**
   Several topics in negative reviews centred on slow or unhelpful service. Investing in staff training focused on communication, conflict resolution, and proactive support could improve guest interactions and reduce service-related dissatisfaction.
3. **Streamline the check-in and room allocation process.**

Issues with incorrect or inconvenient room assignments were common. Improvements here could reduce friction at the start of a guest's stay, improving overall impressions.

4. **Maintain and capitalise on strengths.**
   Positive reviews show that staff friendliness, breakfast, and cleanliness are valued highly. These should be consistently monitored and promoted as differentiators in marketing and guest engagement.

5. **Address noise and sleep quality.**
   Several guests mentioned noise disturbances or uncomfortable beds. Consider room soundproofing, clearer room descriptions, and better bedding options in future refurbishments.

By targeting these issues, the hotel can not only reduce the frequency of negative reviews but also strengthen the elements that drive loyalty and positive word of mouth.