

TOURISM DEVELOPMENT AND REGIONAL DYNAMICS IN THE REPUBLIC OF NORTH MACEDONIA, 2010-2025

Anisa SULIMANI¹, Shkurte LUMA-OSMANI¹, Hatibe DEARI-ZEQIRI², Artina
KAMBERI³

¹ Department of Informatics, Faculty of Natural Sciences and Mathematics, University of Tetova

² Department of International Business, Faculty of Economics, University of Tetova

³ Department of Finance, Faculty of Economics, University of Tetova

Corresponding author e-mail: a.sulimani230113@unite.edu.mk

Abstract

This paper examines tourism dynamics in the Republic of North Macedonia from 2010 to 2025 using official tourism and catering data. The analysis focuses on tourist arrivals, overnight stays, seasonality, regional distribution, source markets, accommodation structure and basic predictive modelling. The dataset was cleaned before analysis, and pre-existing summary rows were excluded to avoid double counting. Missing August and September 2024 records in the working monthly file were imputed with historical monthly averages by region and tourist type, because these months belong to the peak tourist season and their omission would create an artificial decline. The results show long-term growth, a sharp interruption in 2020 and a strong recovery by 2025. Demand remains concentrated in July and August, the Southwest Region dominates overnight stays, and Skopje is one of the main urban tourism centers. Source-market results show the importance of Turkey and neighboring countries, while accommodation results confirm the role of private rentals and individual hotel categories. Decision Tree classification outperformed KNN, while linear regression produced a positive but moderate baseline trend. The findings support more balanced regional planning, stronger off-season tourism products and a diversified market strategy.

Keywords: tourism analytics, North Macedonia, regional tourism, seasonality, accommodation, forecasting, data mining.

1. Introduction

Tourism is more than the movement of visitors from one place to another. It affects employment, local services, cultural promotion, infrastructure and the visibility of regions. Arrivals and overnight stays are therefore two basic indicators of destination performance. Arrivals show visitor volume, while overnight stays show how long demand remains in the destination (United Nations and World Tourism Organization, 2010).

The Republic of North Macedonia includes several tourism profiles in a relatively small territory. Lake tourism, urban visits, spa tourism, rural and nature-based trips and short cross-border flows all appear in the same national tourism system. These forms do not grow at the same pace. A national total can therefore hide important differences between regions, months, origin countries and accommodation categories. For this reason, this article uses official statistical data and separates the analysis into clear thematic parts (State Statistical Office of the Republic of North Macedonia, 2025).

The aim of the paper is to identify patterns that can support interpretation and planning. The analysis follows the data-mining logic of moving from raw data to useful knowledge, but the emphasis is placed on the practical meaning of the results rather than on technical procedures (Fayyad et al., 1996), (Luma-Osmani, et al., 2020).

2. Data and analytical approach

The empirical part is based on tourism and catering data from MAKSTAT, where various field of study data analysis such as Migration (Sali, S. et al., 2025) or Industrial Production (Jashari, P., 2025) can be found. Three data views were used: region and month, country of origin, and accommodation category. Together, these views answer four practical questions: when tourists arrive, where they go, where they come from and which accommodation types they use. Before analysis, the original tables were reorganized into a cleaner structure. Year, month, season and average length of stay were added as analytical variables. Average stay was calculated as overnight stays divided by tourist arrivals. Summary rows that already existed in the original files were not treated as observations because they would distort graphs and models when combined with detailed categories.

A specific data-quality correction was applied to 2024. In the working monthly file, August and September 2024 were missing. Since these two months normally belong to the strongest part of the tourism year, their values were filled with historical averages for the same month, region and tourist type. After this correction, 2024 becomes more comparable with the surrounding years and no longer appears as an artificial collapse caused by missing peak-season records.

Table 1. Datasets and analytical purpose

Dataset	Main content	Purpose in the article
Region/month	Arrivals and overnights by region, month and tourist type	Annual dynamics, seasonality, regional comparison and modelling
Country of origin	Arrivals and overnights by source market	Source-market structure and diversification analysis
Accommodation	Arrivals and overnights by accommodation category	Accommodation structure and overnight distribution

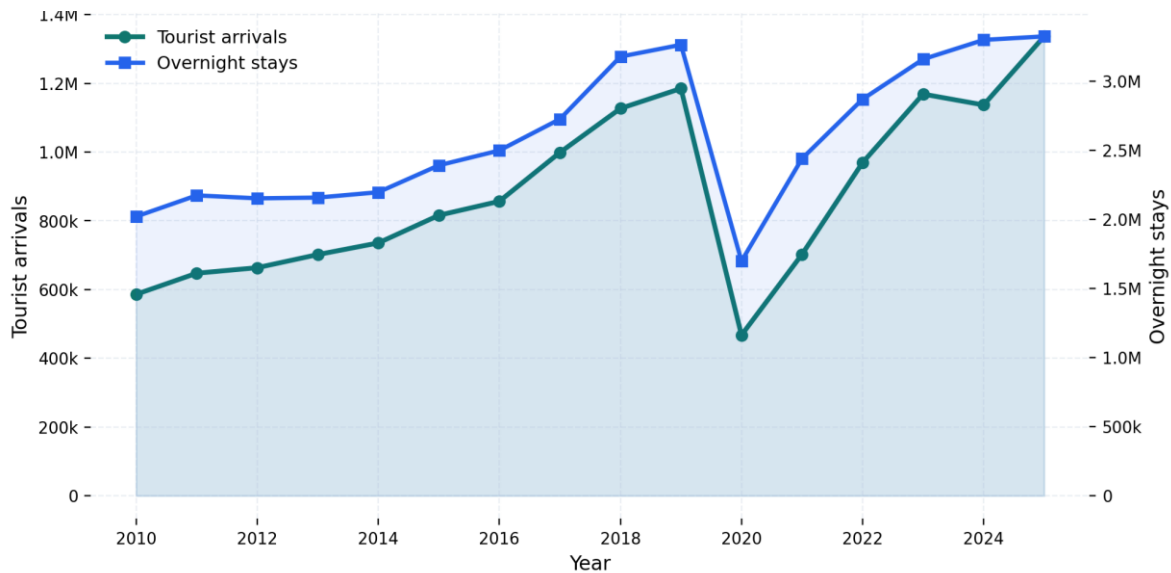
The study combines descriptive statistics, visual comparison, KNN classification, Decision Tree classification and a simple linear-regression forecast. The models are used as supporting tools for interpretation, not as causal explanations.

Table 2. Main indicators used in the analysis

Indicator	Meaning	Analytical role
Tourist arrivals	Number of registered tourists	Measures visitor volume and recovery
Overnight stays	Number of nights spent by tourists	Measures intensity and economic importance
Average stay	Overnights divided by arrivals	Shows whether visits are short-stay or long-stay
Seasonal share	Share of arrivals by season	Identifies peak periods and concentration
Model accuracy	Correct classifications divided by test cases	Evaluates predictive performance

3. Results

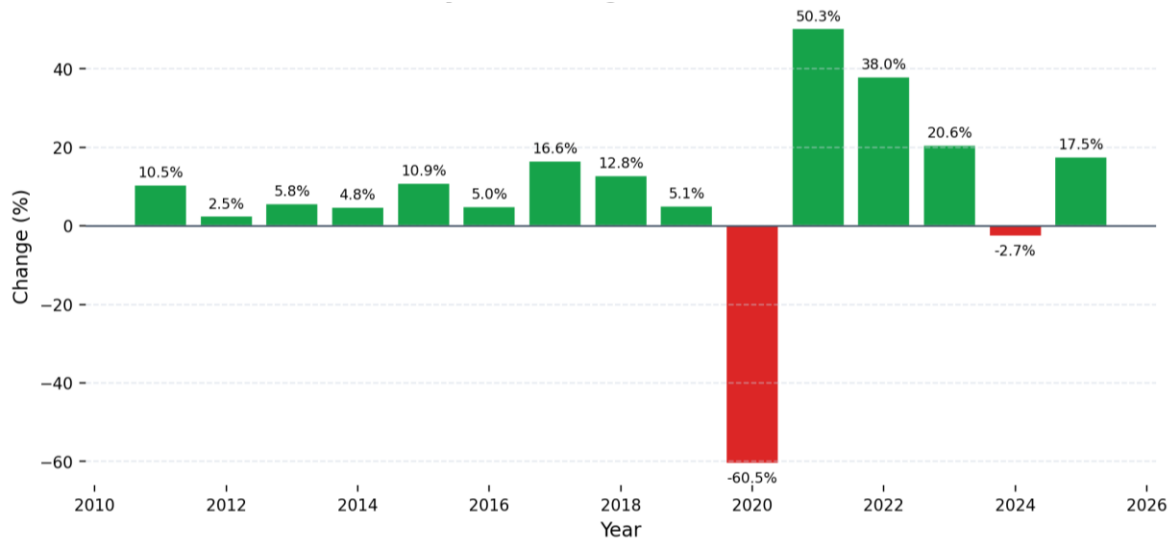
3.1. Annual tourism dynamics: Tourism increased over the analyzed period, but the increase was not linear. Arrivals rose from 586,241 in 2010 to 1,184,963 in 2019. The strongest break occurred in 2020, when arrivals fell to 467,514. After that shock, the sector recovered and reached 1,336,597 tourists in 2025. The corrected 2024 value is 1,137,111 tourists, which places it close to the pre-2025 recovery level rather than indicating a large artificial decline caused by missing summer data.

**Figure 1.** Tourism dynamics in selected benchmark years after the 2024 correction

The year-on-year comparison separates a real shock from a data-coverage issue. The 2020 decline remains the largest change, at about -60.5%. After the imputation of August and September 2024, the 2024 decline is about -2.7%. The following increase in 2025 therefore represents recovery when compared with a corrected previous year.

Table 3. Selected annual indicators after data correction

Year	Tourists	Overnight stays	Interpretation
2010	586,241	2,020,217	Starting point of the series
2019	1,184,963	3,262,398	Pre-pandemic high point
2020	467,514	1,697,535	Strong pandemic decline
2024	1,137,111	3,297,732	Corrected year with imputed August-September values
2025	1,336,597	3,323,210	Highest arrivals in the analyzed period

**Figure 2.** Year-on-year change in tourist arrivals after the correction of missing 2024 peak-season values

3.2. Monthly and seasonal concentration: The monthly pattern confirms strong seasonality. July and August are the most intensive months, with more than 2.3 million and 2.5 million aggregated arrivals respectively. This explains why missing values for August and September 2024 had to be corrected: without them, annual and seasonal results would underestimate the strongest part of the tourism year.

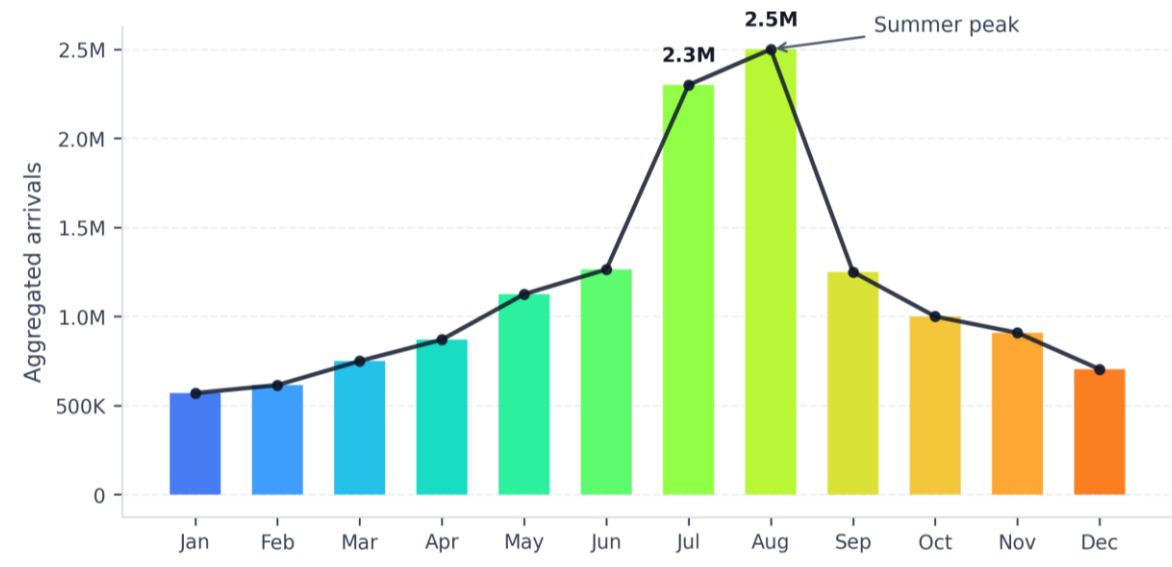


Figure 3. Monthly distribution of tourist arrivals across the analyzed period

Summer accounts for the largest share of arrivals, while winter has the lowest share. This pattern creates pressure on accommodation, transport and local services during peak months, while part of the capacity remains underused outside summer. Planning should therefore strengthen cultural events, rural tourism, spa tourism, conferences and nature-based products in spring and autumn.

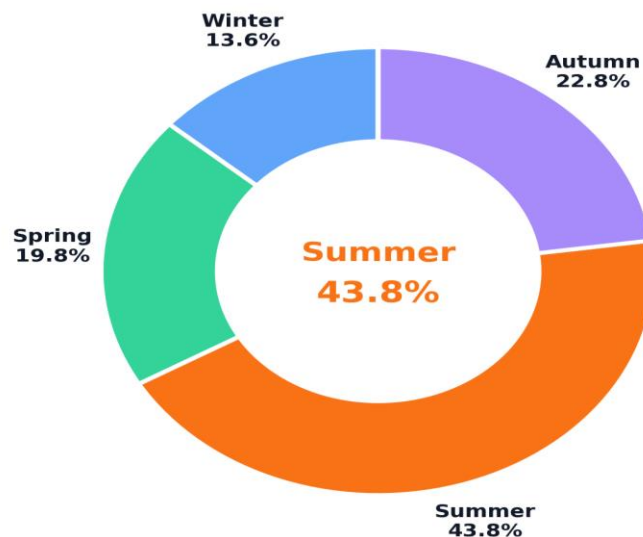


Figure 4. Seasonal concentration of tourist arrivals

3.3. Regional tourism patterns: Tourism growth is spatially uneven. The two strongest centers, Skopje and the Southwest Region, increased considerably between 2010 and 2025. Skopje is mainly an urban, administrative and transit-oriented destination, while the Southwest Region is strongly connected with lake tourism and longer stays. These different profiles require different planning approaches.

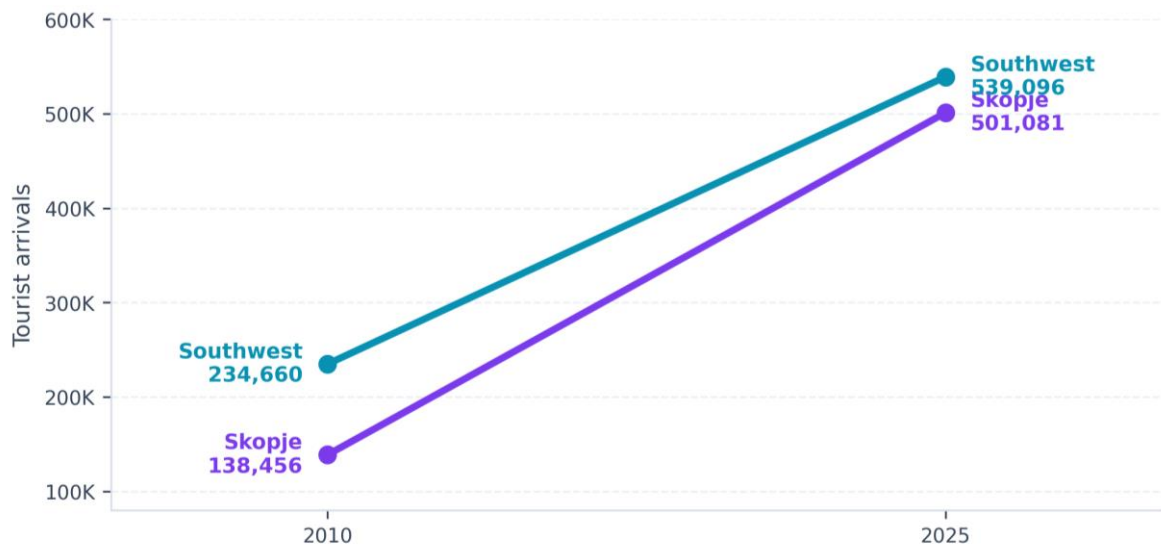


Figure 5. Growth in tourist arrivals in Skopje and the Southwest Region, 2010 and 2025

Overnight stays make the regional concentration clearer. The Southwest Region accounts for more than half of the leading-region overnight structure, followed by Skopje and the Southeast Region. High-demand regions need infrastructure, quality control and sustainability management. Lower-demand regions require stronger visibility, product development and transport connectivity.

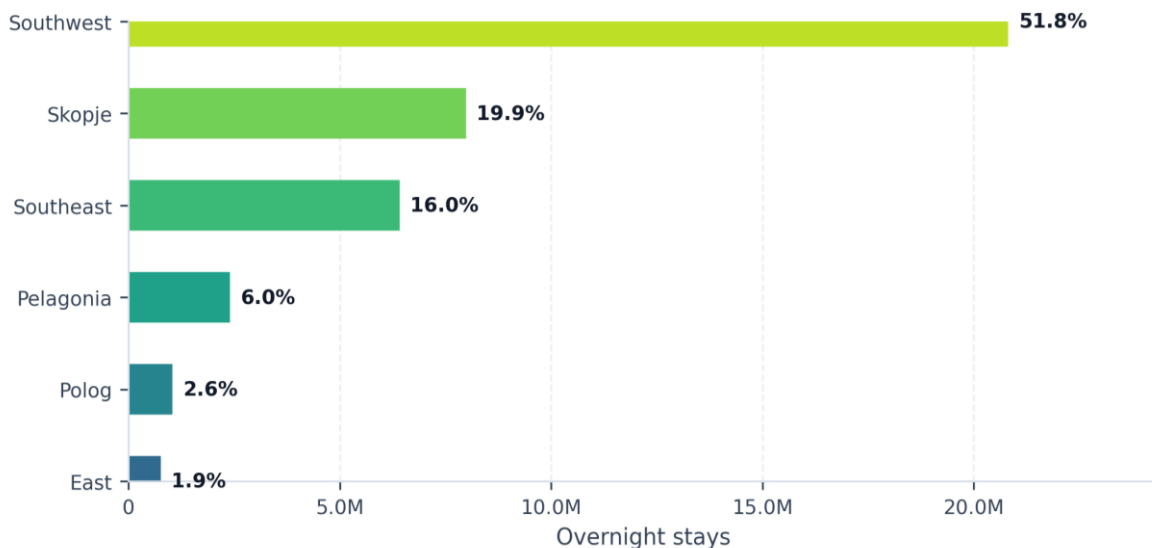


Figure 6. Regional concentration of overnight stays

3.4. Domestic and foreign tourism structure: The balance between domestic and foreign tourism changed noticeably. Domestic tourists represented 55.4% of arrivals in 2010. In 2020, during the disruption of international mobility, the domestic share rose to 74.7%. By 2025, foreign tourists represented 68.5% of arrivals. This confirms a strong recovery of international demand, but it also means that the sector is more exposed to external markets and transport conditions.

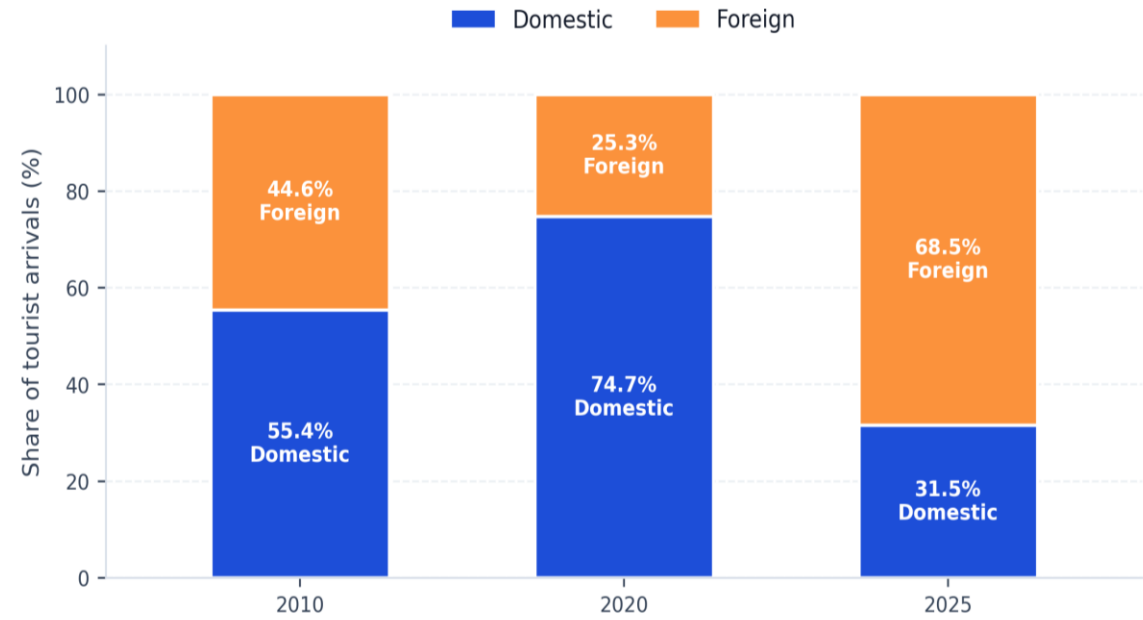


Figure 7. Domestic and foreign tourist structure in selected years

3.5. Source markets and diversification: The source-market results are presented descriptively and in a clearer comparative form. Turkey is the largest source market in the top-ten group, followed by Serbia, Greece and Bulgaria. The strong position of neighboring and regional markets shows that cross-border mobility remains important for North Macedonian tourism. The chart also shows that the structure is concentrated, but not dominated by only one market.

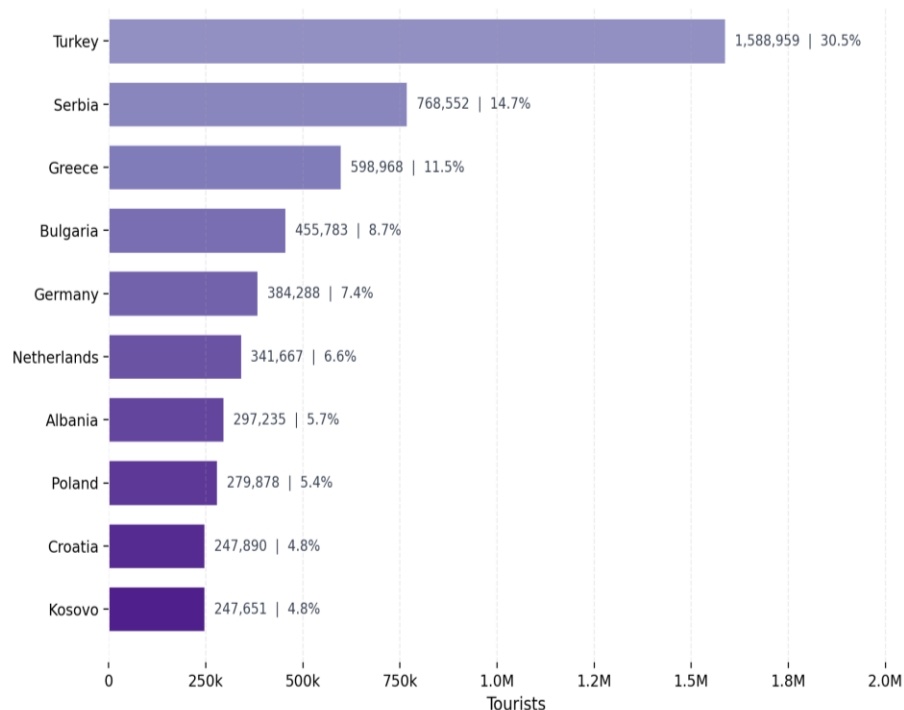


Figure 8. Top ten source markets by tourist arrivals and share within the top-ten group

The main implication is that promotion should maintain the strongest markets but should not depend on only one group of visitors. Diversification toward additional regional and European markets can reduce risk if demand from a large market weakens.

3.6. Accommodation structure: Accommodation analysis links demand with the capacity that receives tourists. To avoid distorted interpretation, total hotel rows and other pre-existing summary rows are not used as observations. The analysis therefore focuses on detailed accommodation categories such as private rentals, hotels by star category, spas, camps and youth facilities. This gives a clearer view of where overnight stays are actually registered.

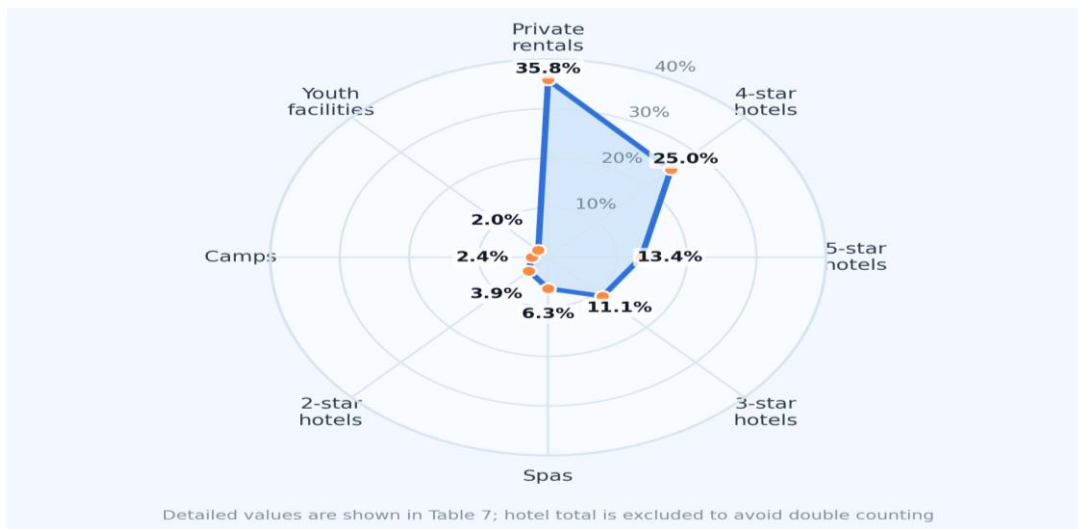


Figure 9. Main accommodation categories by overnight stays, excluding total hotel and other summary rows

The detailed accommodation view shows that private rentals and four-star hotels are the strongest categories by overnight stays. This should be interpreted as a distribution of registered overnight stays, not as a direct measure of profitability. Hotels remain important, but the analysis does not count a hotel total together with its subcategories because that would double count the same structure.

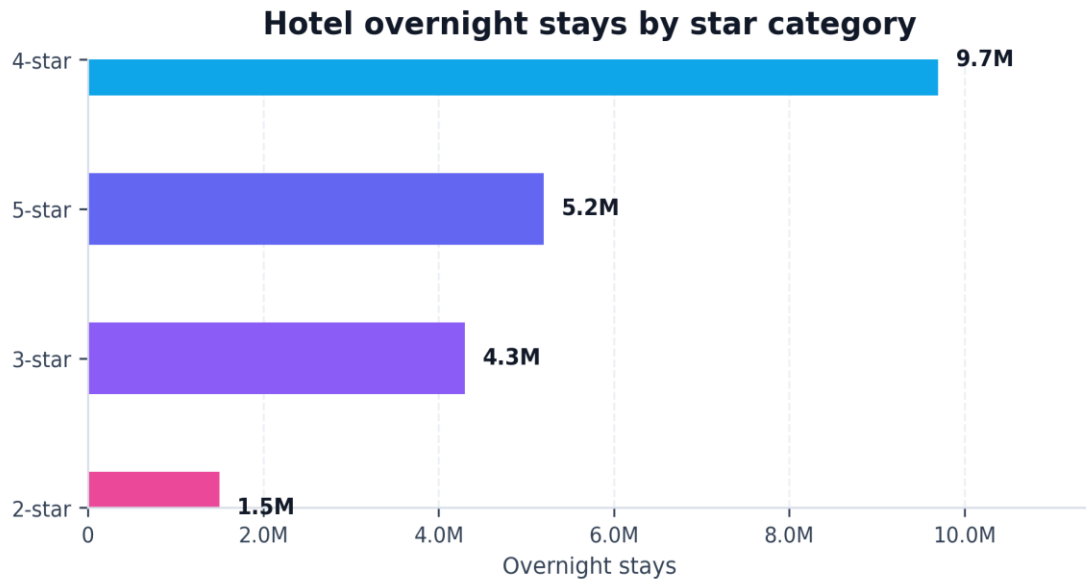


Figure 10. Hotel overnight stays by star category

3.7. Model results and forecast: The modeling results are used only as supporting evidence. KNN and Decision Tree were tested to classify observations into tourism categories. Decision Tree reached the higher accuracy, which means that it separated the categories more clearly in this dataset. This result should not be read as a causal explanation, but as a practical classification result.

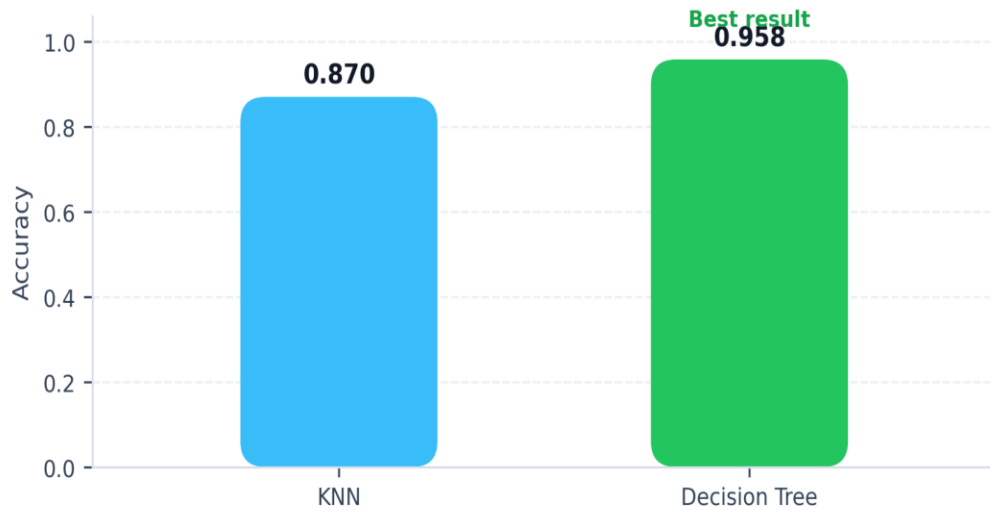
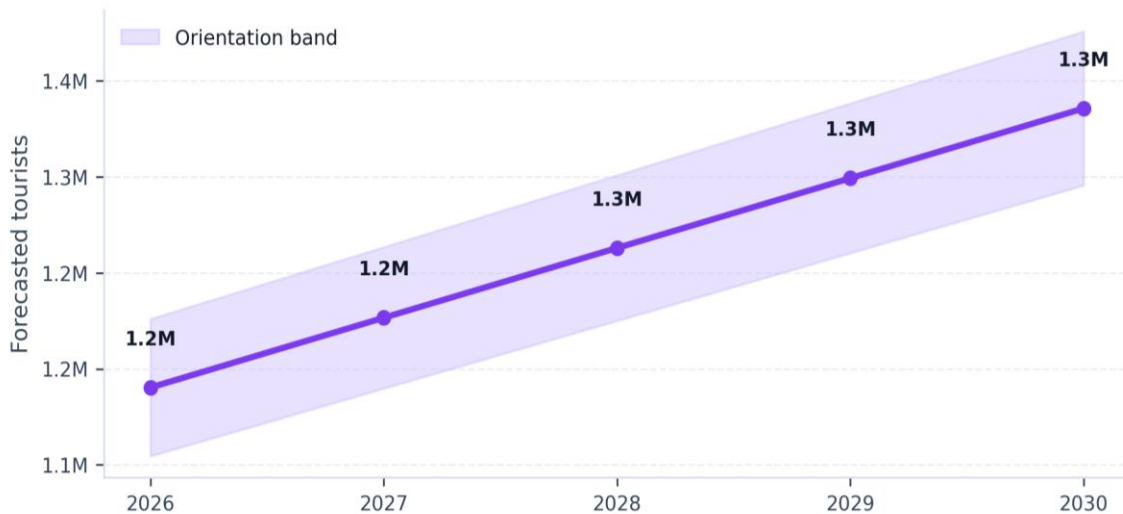


Figure 11. Classification accuracy comparison between KNN and Decision Tree

Table 4. Summary of modeling results

Method	Main result	Interpretation
KNN classification	Accuracy 0.870	Good performance, but weaker than Decision Tree
Decision Tree classification	Accuracy 0.958	Best classification result in this study
Linear regression	$R^2 = 0.458$; positive slope	Positive baseline trend, but only moderate explanatory power

Linear regression was used as a simple baseline forecast. The positive direction indicates continuing recovery, but the R^2 value of 0.458 shows that year alone cannot explain tourism dynamics. Future forecasts should include transport accessibility, accommodation capacity, prices, events and macroeconomic indicators.

**Figure 12.** Baseline linear-regression forecast for 2026-2030

4. Discussion and policy implications

The findings show a tourism sector that has recovered strongly but remains uneven across time and space. The 2020 decline reflects a real disruption, whereas the corrected 2024 value shows the importance of data quality. Missing peak-season records can create a misleading picture of decline; therefore, data cleaning is a necessary part of tourism planning.

Spatial concentration is the second major issue. The Southwest Region dominates overnight stays, while Skopje is a major urban tourism centre. These two areas need different strategies: high-demand areas require infrastructure, sustainability measures and service-quality management, while lower-demand regions need better product development, visibility, events and transport links.

Seasonality is the third issue. Summer, especially July and August, remains the centre of tourism demand. A stronger off-season offer based on cultural events, rural tourism, spa tourism, conferences and nature-based activities would help distribute demand more evenly throughout the year.

Accommodation results show that both hotels and private rentals matter. However, hotel totals should not be mixed with hotel subcategories in the same analysis. Hotels are more formal and easier to monitor, while private rentals can expand capacity quickly but require consistent registration and quality standards. A realistic tourism strategy should include both forms of accommodation.

5. Conclusions

This paper analyzed tourism dynamics in the Republic of North Macedonia from 2010 to 2025 using descriptive statistics, visual analysis and selected predictive methods. The results confirm long-term growth, a strong decline in 2020 and a recovery by 2025. The study also corrects the interpretation of 2024 by imputing missing August and September values with historical averages. The main findings are clear. July and August are the strongest months. The Southwest Region dominates overnight stays. Skopje is a key urban tourism center. Turkey and neighboring countries are important source markets. Private rentals and individual hotel categories form the core of accommodation capacity, but summary totals are excluded from the category-level analysis. Decision Tree classification achieved the highest model accuracy, while linear regression indicated a positive but only moderately explained trend.

Future research could improve the forecast by adding transport accessibility, accommodation capacity, prices, major events, macroeconomic variables and more advanced time-series models. Such additions would make the findings even more useful for tourism planning and regional development policy.

References

- [1]. Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, Vol. 46, No. 3, pp. 175-185.
- [2]. Jashari, P., Luma-Osmani, S., Memeti, A., & Ajvazi, G. (2025). Analysis of Trends and Anomalies in the Industrial Production Index in North Macedonia. *Journal of Natural Sciences and Mathematics of UT*, 10(19-20), 363-369.
- [3]. Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. 1984. Classification and Regression Trees. *Wadsworth International Group*, Belmont, California.
- [4]. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. 1996. From data mining to knowledge discovery in databases. *AI Magazine*, Vol. 17, No. 3, pp. 37-54.
- [5]. Pedregosa, F. et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830.
- [6]. Luma-Osmani, S., Ismaili, F., Zenuni, X., Raufi, B.: A systematic literature review in causal association rules mining. In: 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0048–0054 (2020). <https://doi.org/10.1109/IEMCON51383.2020.9284908>

- [7]. State Statistical Office of the Republic of North Macedonia. 2025. MAKSTAT Database: Tourism and Catering datasets, 2010-2025. Accessed 7 July 2026.
- [8]. United Nations and World Tourism Organization. 2010. *International Recommendations for Tourism Statistics* 2008. United Nations, New York.
- [9]. Sali, S., Luma-Osmani, S., Imeri, F., & Memeti, E. (2025). Data-Driven Insights into Migration: Identifying Key Drivers of Population Decline in North Macedonia. *Journal of Natural Sciences and Mathematics of UT*, 10(19-20), 401-405.