**Analyzing Visitor Arrivals in Qatar: Statistical Analysis**

**Name: Mohammed Aly**

**Student ID: 101825**

**Introduction**

Accurate forecasting of visitor arrivals is essential for planning and optimizing tourism infrastructure in Qatar. This report explores various data analysis techniques, including data cleaning, visualization, correlation analysis, anomaly detection, and statistical tests, to provide a comprehensive overview of visitor arrivals via Air, Land, and Sea modes.

The report aims to analyze visitor arrival data by evaluating distributions, identifying outliers, performing correlation analysis, detecting anomalies using statistical tests, and assessing the growth rate over time. The results will guide tourism strategies and infrastructure development.

The objectives of this report are:

* To clean and preprocess the data, identifying missing values and duplicates.
* To visualize and analyze the distribution of visitor arrivals.
* To perform correlation analysis and detect anomalies in the data.
* To investigate seasonal trends and perform statistical tests, including t-tests and Mann-Whitney U tests, to identify any significant anomalies.
* To calculate the yearly growth rate of visitor arrivals.

**Procedure**

The data for this report was collected from the Qatar Monthly Statistics on Visitor Arrivals dataset, containing information on visitor arrivals by Air, Land, and Sea, along with the total arrivals per month.

* **Data Cleaning:** Python libraries such as pandas were used to handle missing values and duplicates.
* **Data Visualization:** matplotlib, seaborn for plotting histograms, boxplots, and trend analysis.
* **Correlation Analysis:** Pearson correlation was used to determine relationships between variables.
* **Anomaly Detection:** Z-scores and boxplots were employed to identify potential outliers.
* **Statistical Testing:** T-tests and Mann-Whitney U tests were used to compare specific months and detect significant differences in visitor arrival patterns.
* **Growth Rate Calculation:** Percentage change from the previous year was calculated to assess the yearly growth rate.

**Findings and Analysis:**

The data analysis of visitor arrivals by mode of entry (Air, Land, Sea) and total visitor arrivals reveals several important trends and statistical observations:

**A graph of different colored lines

AI-generated content may be incorrect.**

Figure 1: Visitor Arrivals by Mode of Entry (Air, Land, Sea) and Total over Time

The **line graph of visitor arrivals by mode of entry (Air, Land, and Sea) and total arrivals over time** clearly illustrates the fluctuating trends across the years. A significant spike in Air Arrivals can be observed in 2019, which aligns with a noticeable rise in total visitor arrivals. Meanwhile, Sea and Land Arrivals remain relatively stable but show a sharp decrease around 2020, reflecting global travel restrictions during the COVID-19 pandemic. This shift highlights the dominant role of Air Arrivals in visitor numbers and the vulnerability of Land and Sea travel during such events.

A graph of a number of months

AI-generated content may be incorrect.

Figure 2: Average Monthly Visitor Arrivals

The **monthly average visitor arrivals** chart further emphasizes seasonal patterns, with a steady increase in Air Arrivals and Total Visitor Arrivals during the later months of the year. **Air Arrivals** show the most significant upward trend, indicating increased international travel during peak months, whereas **Sea and Land Arrivals** exhibit lower fluctuations, suggesting more localized or seasonal travel.

A graph with a line going up

AI-generated content may be incorrect.

Figure 4: Yearly Growth Rate

The **yearly growth rate of total visitor arrivals** shows considerable volatility, with a sharp rise in 2022. The massive growth spike can be attributed to the gradual recovery of travel post-pandemic. In contrast, the negative growth rates for 2020 and 2021 reflect the drastic decline in international and regional travel due to COVID-19 restrictions. This year-on-year comparison provides a clear understanding of how global events and policies can impact visitor arrivals, with rapid recovery observed once restrictions eased.

**Recommendations**

1. **Focus on Enhancing Air Travel Infrastructure:** Given the strong correlation between Air Arrivals and total visitor numbers, it is recommended that significant investments be made to enhance air travel infrastructure.
2. **Seasonal Promotions for Land and Sea Arrivals**: To boost the relatively lower Land and Sea Arrivals, seasonal promotions and packages should be introduced. Partnering with local travel agencies and providing special incentives during off-peak seasons could increase the share of visitors traveling by land and sea.

**Conclusion:**

The analysis of visitor arrivals data has provided valuable insights into the trends and patterns of tourism, emphasizing the dominance of Air Arrivals in driving the overall increase in visitor numbers. Notably, the recovery of the tourism sector post-pandemic in 2022 demonstrates the resilience of air travel. However, Land and Sea Arrivals have lagged behind and require more targeted attention. The **yearly growth rate** data provides evidence that the tourism industry is on a recovery trajectory, and with proper management, the growth rate can continue to be positive. However, seasonal fluctuations remain a challenge, requiring enhanced planning and preparation.

**References**

* Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning* (3rd ed.). Packt Publishing.
* Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python*. JMLR, 12, 2825-2830.
* James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.