VISITORS TO KENYA

# PROBLEM STATEMENT

Owing to the fact that there are many people who benefit from visitors and visits in the country. It would be prudent to understand the trends related to these visitors coming in and out of the country as it would be beneficial to many.

# OBJECTIVES AND GOALS

The objective of this project is to get a summary of the trend related to the visits in Kenya and have data visualizations of the same.

Some of the questions to be answered are:

1. What is the trend of visitors over the years?
2. Which purpose is most dominantly used by visitors? And what has its trend been?
3. Do the quarters affect the visitors, and if yes how?
4. What is the trend of the quarters over the years?

# PROJECT PLAN

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PLAN | DURATION | RESOURCES | INPUT | OUTPUT | DEPENDENCIES |
| Problem statement | 15 min | N/A | Local Observations | Problem statement | N/A |
| Objectives and goal setting | 15 min | N/A | Problem statement | Project goals and objectives | Problem statement |
| Data collection | 5 min | N/A | N/A | Working data | Problem statement |
| Data preparation and quality | 15 min | Python libraries | Working data | Data quality report | Collected data |
| Analysis | 1 hr 30 min | Python libraries and Vscode | Prepared data | Analysis report and data visualization | Tools and libraries used |
| Conclusion | 15 min | Word document | Project summary | Project findings | Project summary |
| Recommendation and next step | 15 min | Word document | conclusions | Recommendations from the project | Analytical findings |

# DATA COLLECTION

The data was collected from a csv file known as visitors\_to\_kenya.csv

# DATA PREPARATION AND QUALITY

The initial steps to preparing the data were to open a jupyter note book and to import python libraries then load the csv file in our working environment.

The data set had 832 entries and 8 columns. The columns of the data set were as follows:

|  |  |  |
| --- | --- | --- |
| COLUMN INDEX | COLUMN NAME | DATA TYPE |
| 0 | Arrival\_or\_Depature | Object |
| 1 | Date\_End\_of\_Quater | Object |
| 2 | Year\_Text | Object |
| 3 | No\_of\_Visitors\_to\_Kenya | Int64 |
| 4 | Visitors\_by\_Purpose | object |
| 5 | Period\_Quater | Object |
| 6 | Result\_Status | object |
| 7 | OBECTID | Int64 |

There were no recorded null values as the information showed 832 non-null values in every column.

# DATA CLEANING

Considering the nature of the Year\_Text column it was necessary to convert it to integer. First, the string characteristics; \* and ‘ at the end of some of the entries were removed, making it possible for conversion of the column to numeric.

Then the columns were renamed for easy manipulation. The new columns were as follows: move, date, Year\_Text, num, purpose, Q, status, id, year, x. From there I picked the column that I’d need for the analysis.

The actual conversion of the year column to numeric was done.

The purpose column had long strings that would make manipulation more tedious than if they are one word. The values in this column were changed as follows; ‘Visitors on Holiday’ to ‘holiday’, ‘Visitors on Business’ to ‘business’ and ‘Visitors in Transit’ to ‘transit’.

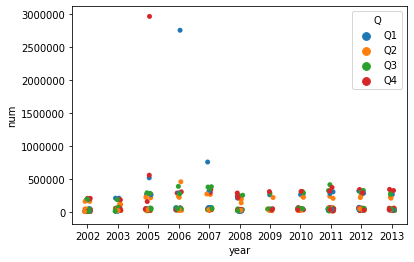
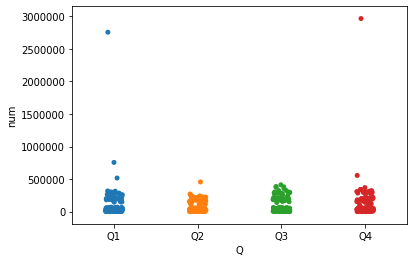
The analysis would focus only on the actual entries therefor rendering the ‘preliminary’ entries useless to the project. Also, we’d analyze the data from 2002, hence the entries from the years before that were also useless to the project. Due to that, only the entries whose status was actual and year above 2001 were picked. This reduced our working data frame from having 832 entries to 336.

The next step was to check for duplicate values to which there were none.

## OUTLIERS

In relation to quarters;

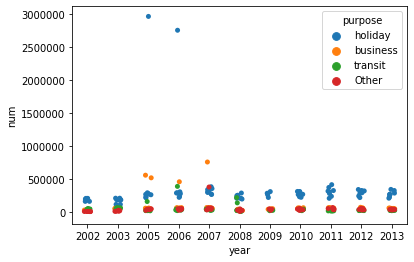
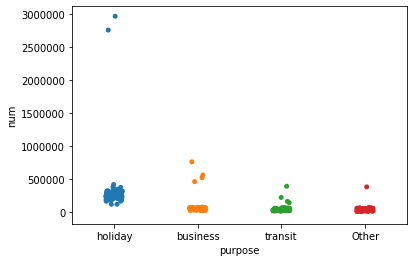
There seemed to be an outlier in the year 2005 Q4 and 20006 Q1, the num variable was above 2,500,000 while most of the other values were below 500,000.



.

In relation to purpose;

There seemed to be an outlier in the year 2005 and 2006 in relation to holiday while the others were below 500,000. The other ones that were above 500,00 were around 2005, 20006 and 2007 in relation to business. The holiday purpose had values above 2,500,00 while the business purpose had values above 500,000.

# ANALYSIS

## UNIVARIATE

The description of the num variable was as follows:

Mean 🡪 115,643

Minimum 🡪 3,800

25% 🡪 26,950

50% 🡪 44,850

75% 🡪 180,900

Maximum 🡪 2,966,000

Standard deviation 🡪 167,237

std 🡪 242,241.2

Skew 🡪 11.414670075091141

Kurtosis 🡪 183.64413696629836

## BIVARIATE

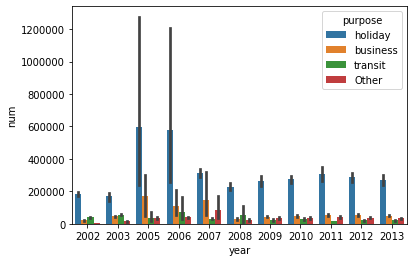
The total number of visitors per year was an increasing trend from 2002 and only sharply dropped from 2006 to 2009 and started slowly increasing again over the years. The peak was at 2005 with 6,714,300 followed by 2006 with 6,417,200 and 2007 with 4,607,400.

The total number of visitors per quarter in that period drops sharply from Q1 to Q2 and the starts rising up to Q4. The highest quarter is Q4 with 11,716,000 followed by Q1 11,254,300 and Q3 with 8,971,100. Lastly Q2 had 6,914,700.

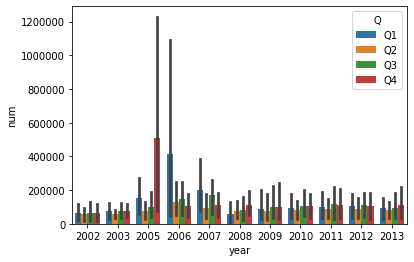
The total number of visitors by purposes is highest for holiday with 26,734,400 followed by business with 5,970,300 and transit with 3,176,500. Lastly other purposes had 2.974,900.

## MULTIVARIATE

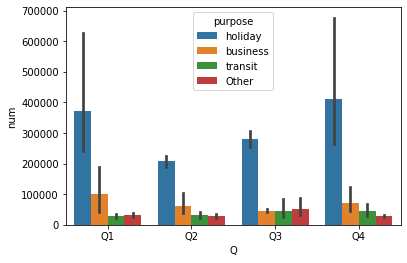
1. Year per purpose:



1. Year per quarter:



1. Quarter per purpose:



# CONCLUSION

Holiday is the most used purpose for travel to Kenya, over the years and in every quarter.

Business is the second mostly used purpose of travel, with transit and others following.

The quarterly trend is most high in Q4 and Q1, followed by Q3 and lastly Q2.

Over the years 2005, 2006 and 2007 have recorded the most travel entries in the country.

# RECCOMMENDATION AND NEXT STEP

Further research; other than the one available on the data will be done.