GovHack 2024 Project – Holiday Home NT

Event: Darwin

Region: Northern Territory

Team Name: Grey - 2569

Team Members:

1. Mohammad Jobaidul Hoque

2. Synthia Islam

YouTube Link: https://youtu.be/EbOU7jVau9U

GitHub Link: https://github.com/synthia26/GovHack2024

Table of Contents

Table of Figures

Problem Statement:

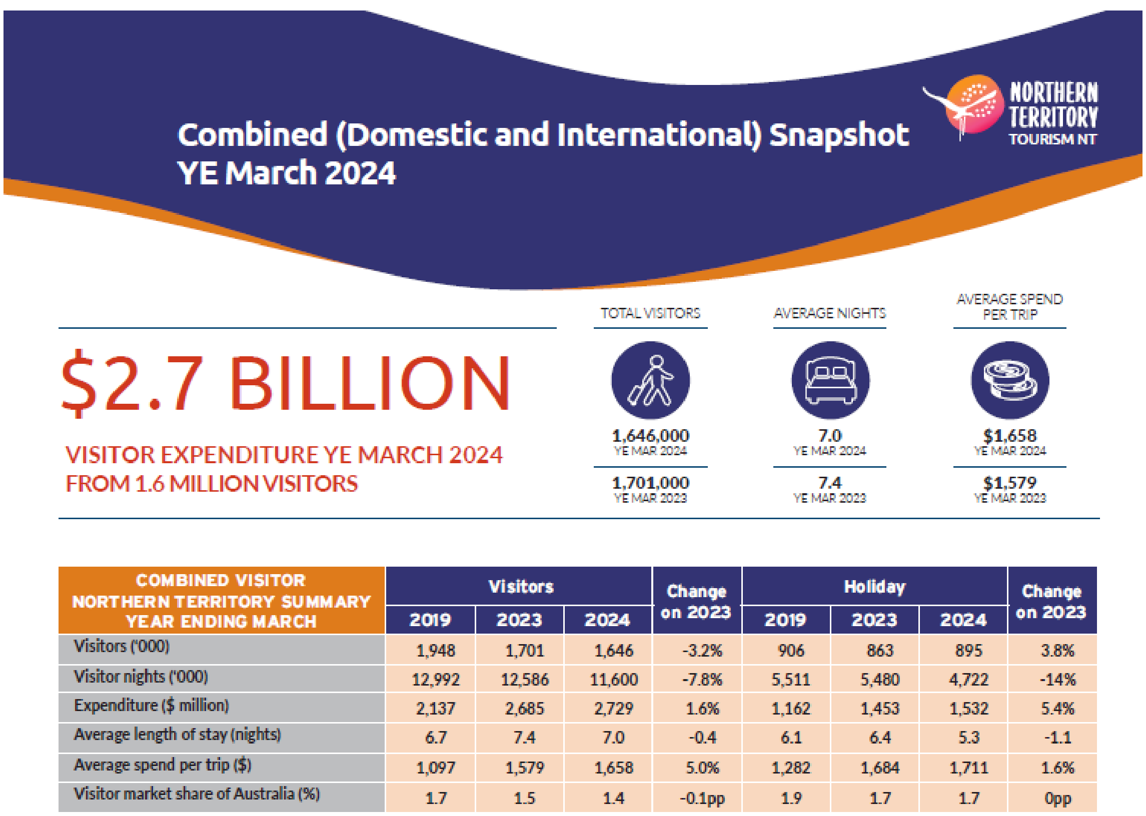
Our analysis of the tourism datasets for the Northern Territory (NT) has revealed several challenges:

- **Limited Services in Remote Areas**: There is a noticeable lack of tourism-related services in remote areas, especially those focused on promoting local cultures and traditions.

- **Insufficient Accommodation Options**: The availability of accommodations across the NT is limited, particularly in remote regions, making it difficult to attract and accommodate visitors.

- **Lack of Engaging Tourism Activities with Indigenous Communities:** Tourists in NT often miss authentic cultural experiences due to the lack of engaging activities with indigenous communities. Without immersive opportunities like staying with local families or participating in traditional practices, visitors may feel disconnected from the region's rich cultural heritage.

Due to these challenges, the number of visitors is decreasing in the Northern Territory, which is directly impacting the tourism sector of the NT.


Combined (Domestic and International) Snapshot
YE March 2024

NORTHERN TERRITORY
TOURISM NT

$2.7 BILLION
VISITOR EXPENDITURE YE MARCH 2024
FROM 1.6 MILLION VISITORS

TOTAL VISITORS
1,646,000
YE MAR 2024
1,701,000
YE MAR 2023

AVERAGE NIGHTS
7.0
YE MAR 2024
7.4
YE MAR 2023

AVERAGE SPEND PER TRIP
$1,658
YE MAR 2024
$1,579
YE MAR 2023

COMBINED VISITOR NORTHERN TERRITORY SUMMARY YEAR ENDING MARCH	Visitors			Change on 2023	Holiday			Change on 2023
	2019	2023	2024		2019	2023	2024	
Visitors ('000)	1,948	1,701	1,646	-3.2%	906	863	895	3.8%
Visitor nights ('000)	12,992	12,586	11,600	-7.8%	5,511	5,480	4,722	-14%
Expenditure ($ million)	2,137	2,685	2,729	1.6%	1,162	1,453	1,532	5.4%
Average length of stay (nights)	6.7	7.4	7.0	-0.4	6.1	6.4	5.3	-1.1
Average spend per trip ($)	1,097	1,579	1,658	5.0%	1,282	1,684	1,711	1.6%
Visitor market share of Australia (%)	1.7	1.5	1.4	-0.1pp	1.9	1.7	1.7	0pp

Figure 1: Overall NT Visitors Report from NT Tourism
(Source: https://www.tourismnt.com.au/research-strategies/research/latest-visitor-data)

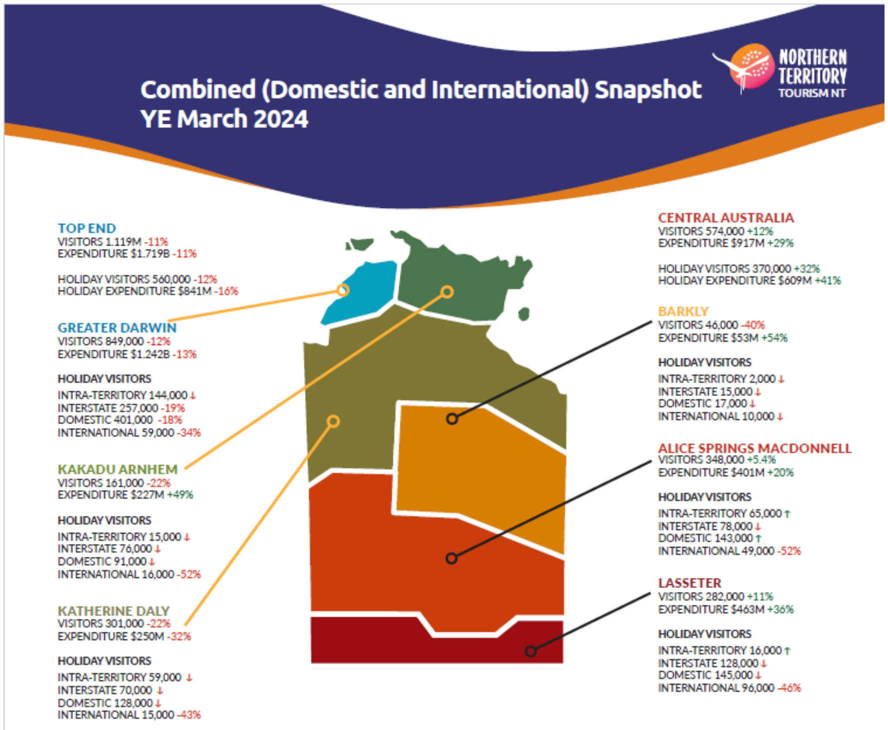


Figure 2: NT Visitors Based on Locations Report From NT Tourism (Source: https://www.tourismnt.com.au/research-strategies/research/latest-visitor-data)

Proposed Solution:

To tackle these issues, our primary solution is the introduction of **Holiday Home NT** as a unique and immersive way for tourists to experience the culture and lifestyle of the Northern Territory's local communities. Additionally, we propose using **an automated data prediction model** to guide and support the promotion and strategic implementation of these holiday homes.

1. Holiday Home Concept (Main Idea):

- This initiative will allow tourists to stay with local families in their homes, allowing them to fully immerse themselves in the traditional Aboriginal lifestyle.
- Tourists will live like locals, sharing meals, engaging in daily activities, and experiencing adventures such as fishing, hunting, and other culturally significant practices.
- By offering visitors a unique and more authentic tourism experience through the Holiday Home concept that promotes local culture and heritage, we aim to fill the gap in accommodation and cultural engagement as well as provide economic opportunities for indigenous communities.

2. Automated Visitor Prediction Model Using Machine Learning (Supporting Idea):

- To ensure the success and sustainability of the holiday home initiative, we will build a machine learning-based model that predicts future visitor trends, including the total number of visitors and the specific locations they are likely to visit.
- These predictions will enable the government and local businesses to identify which areas should be prioritized for promoting holiday homes, particularly in remote regions that are expected to become tourist hotspots in the near future.
- By leveraging these data-driven insights, resources can be allocated more effectively and targeted marketing strategies can be implemented to maximize the impact of the holiday home concept.
  - **Example:**

    i. Future Prediction of Visitors of 2025 for Overall Northern Territory:

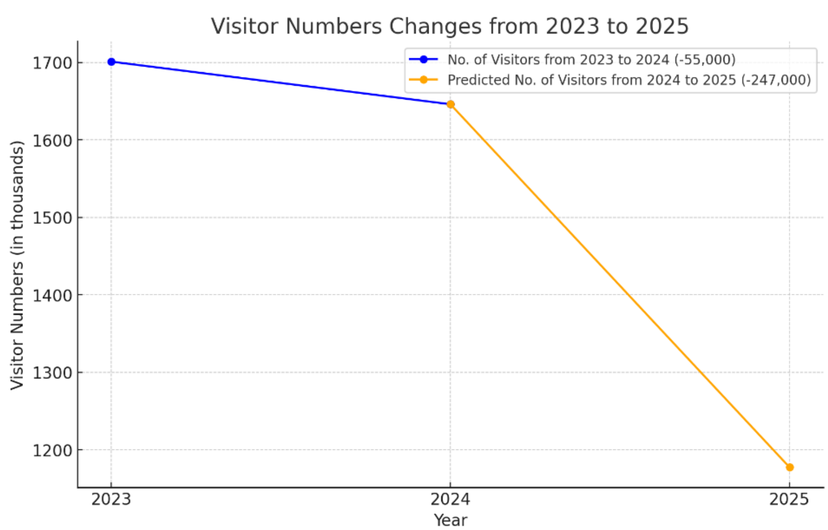


Figure 3: NT Tourist Prediction from Machine Learning Prediction Model

ii. Future Prediction of Visitors of 2025 Based on Different Locations of Northern Territory:

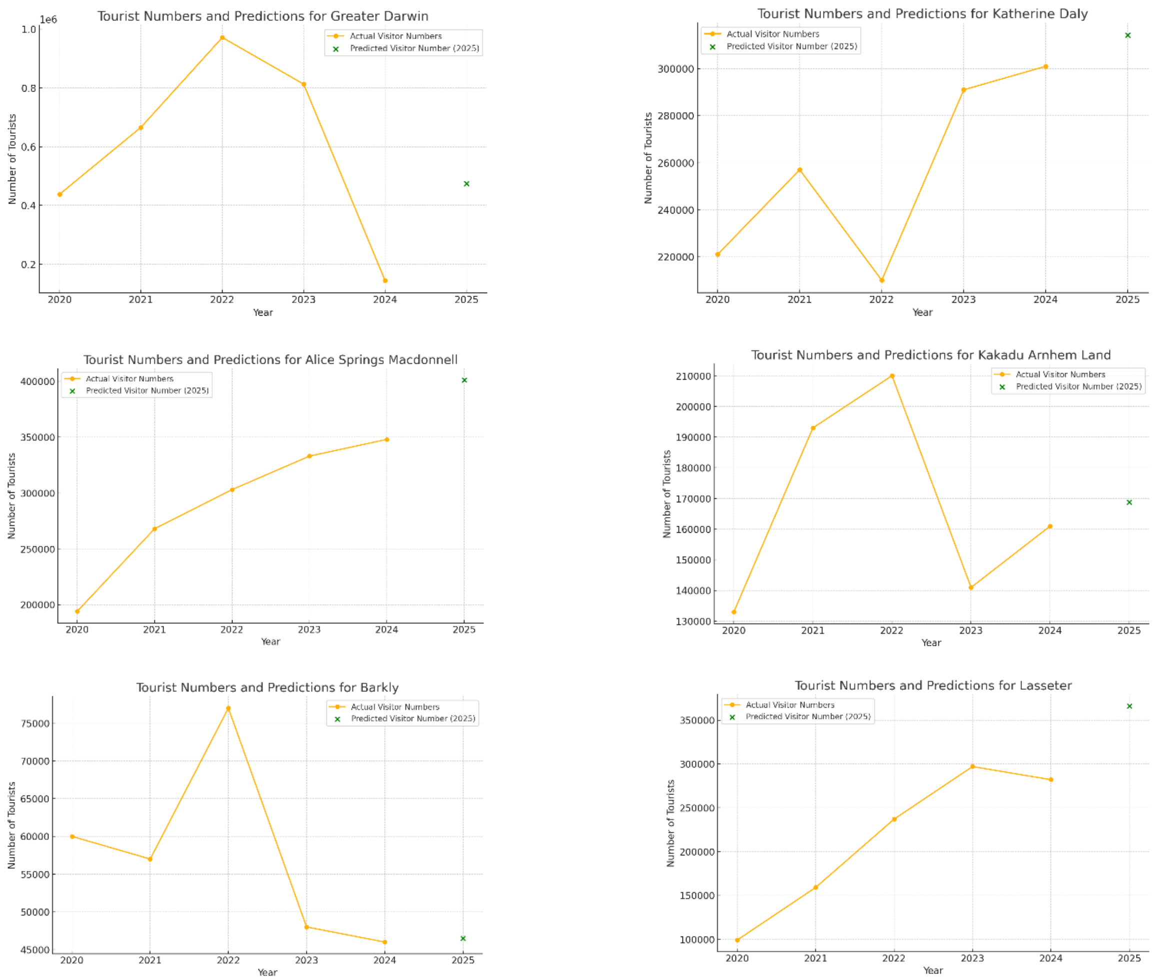


Figure 4: NT Tourist Prediction Based on Locations from Machine Learning Prediction Model

Challenges:

1. **Data Organization for Prediction**: The dataset presented unstructured and diverse information, making it difficult to directly implement machine learning algorithms for visitor prediction. Organizing and cleaning the data to make it suitable for machine learning models was one of the primary challenges.
2. **Idea Generation and Concept Validation**: Developing an innovative tourism concept that promotes local culture while attracting visitors was another challenge. We had to ensure that the concept of holiday homes was not only appealing but also culturally sensitive and economically viable.
3. **Cultural Integration**: Ensuring that the holiday home concept respectfully integrates the indigenous lifestyle and traditions without causing disruption to the local communities was an important aspect to consider.

Overcoming Challenges:

1. **Data Organization**:
   - We conducted a series of brainstorming sessions to critically assess the dataset. After thoroughly analyzing the data and consulting research papers and tourism trends, we managed to clean and structure the dataset.
   - By integrating machine learning tools, we sorted the data into useful formats for our prediction model, ensuring accurate forecasting of visitor numbers and locations of interest.
2. **Collaborative Idea Development**:
   - Working closely with team members and consulting external resources allowed us to refine the holiday home concept. We looked at successful cultural tourism models around the world and adapted those ideas to the unique context of NT, ensuring cultural appropriateness.
3. **Community Engagement**:
   - By involving local communities early in the planning process and promoting the economic and cultural benefits of the holiday home initiative, we were able to gain support from local indigenous leaders. We also plan to provide training and resources to local families, enabling them to participate fully.

4. **Cultural Sensitivity**:
    - To ensure the holiday home initiative remains culturally respectful, we proposed creating guidelines for tourists to follow during their stay, ensuring they respect local customs and practices. This will help protect the local way of life while still offering tourists an authentic experience.

Final Outcome:

1. **Cultural Preservation**: The holiday home initiative will not only promote indigenous culture but also preserve it by making it an integral part of NT's tourism identity.
2. **Unique and Authentic Tourism**: Tourists will experience a one-of-a-kind, immersive adventure that goes beyond the typical commercial tourism offerings. This will position NT as a unique destination for cultural tourism.
3. **Strengthening Local Economies**: Local families and communities will gain financial benefits by participating in the holiday home program, generating income directly from tourism without large-scale infrastructure investments.
4. **Tourism Growth and Economic Impact**: The initiative will contribute to the growth of the NT's tourism industry, bringing in more visitors, increasing tourism revenue, and boosting NT's contribution to Australia's GDP.
5. **Empowerment of Indigenous Communities**: The project will empower indigenous communities by providing them with new opportunities for economic development and cultural exchange, giving them financial independence and control over how their culture is presented.
6. **Sustainable Tourism Practices**: Tourists staying in local homes will experience sustainable practices firsthand, such as traditional ways of living in harmony with nature, reducing waste, and utilizing natural resources responsibly.
7. **Reduced Environmental Impact**: Since the holiday homes utilize existing community resources (local homes and infrastructure), the need for constructing new hotels or large-scale tourism facilities is reduced, leading to less environmental disruption and lower carbon emissions.
8. **Promotion of Eco-Friendly Activities**: Tourists will engage in low-impact activities like fishing, hunting, and learning traditional practices, which are more environmentally friendly compared to conventional tourism activities. These activities will also teach visitors the importance of conservation and sustainable living.

Feasibility:

1. **Cost-Effective and Scalable Machine Learning Model**: The machine learning model is easy to build, cost-effective to maintain, and scalable, allowing it to adapt to new data and predict future tourism trends efficiently.
2. **Resource Efficiency**: By leveraging the resources and infrastructure already present in indigenous communities, the initiative ensures that the tourism experience is resource-efficient, reducing the need for new construction and minimizing energy and water consumption
3. **Ready Local Population Base**: With a significant indigenous population in NT, there is an existing base of communities that can actively participate in the holiday home initiative, ensuring the project can be implemented quickly.
4. **Abundant Natural and Cultural Resources**: NT's unique landscapes and rich indigenous culture create the perfect setting for a tourism model that emphasizes adventure, local traditions, and cultural immersion.
5. **Simple Implementation**: The holiday home concept requires minimal upfront investment and is easy to understand and implement. It leverages existing community resources—local homes and indigenous lifestyles—to offer tourists a distinctive experience.
6. **Alignment with Global Sustainability Goals**: The project supports global sustainability goals by promoting responsible tourism, reducing carbon emissions, and fostering an environmentally conscious approach to travel and accommodation.

Appendix:

1. For Overall NT Tourist Prediction:

i. Dataset: We have created datasets from the source of NT state report of tourism on visitors from 2019-2024(March). All links are given in the reference section.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
vis-dom	vis-int	vis-night-d	vis-night-ir	exp-dom	exp-int	avg_length	avg_length	avg-spend-	avg-spendi	vis-mkt-sh	vis-mkt-sh	purpose-h	purpose-fr	purpose-b	purpose-o	purpose-h	purpose-fr	purpose-b	purpose-o	year
1651000	297000	9663000	3330000	1664000	473000	5.9	11.2	1008000	89000	1.5	3.5	689000	201000	718000	106000	217000	19750	6900	22000	2019
1581000	120000	9381000	3205000	2417000	268000	5.9	18.5	1529000	50000	1.4	3.2	847000	201000	505000	97000	17000	21000	7500	23000	2023
1447000	199000	7056000	4544000	2336000	393000	4.9	22.8	1615000	43000	1.3	2.8	743000	187000	453000	123000	152000	24000	9000	24000	2024

Table 1: Dataset of Overall NT Visitors

ii. Python Code for Machine Learning Model:

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
!pip install -q hvplot
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 155.4/155.4 kB 1.5 MB/s eta 0:00:00
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import hvplot.pandas
from scipy import stats

%matplotlib inline
sns.set_style("whitegrid")
plt.style.use("fivethirtyeight")
```

```
# Load the Tourist dataset
data = pd.read_csv('/content/drive/MyDrive/Synthia-Colab-notebooks/Govhack-data/Tourist-Dataset.csv')
data
```

	vis-dom	vis-int	vis-night-dom	vis-night-int	exp-dom	exp-int	avg_length-stay-dom	avg_length-stay-int	avg-spend-trip-dom	avg-spending-trip-int	...	vis-mkt-share-int	purpose-holiday-dom	purpose-fnf-dom	purpose-business-dom	purpose-others-dom	purpose-holiday-int	purpose-fnf-int
0	1651000	297000	9663000	3330000	1664000	473000	5.9	11.2	1008000	89000	...	3.5	689000	201000	718000	106000	217000	19750
1	1581000	120000	9381000	3205000	2417000	268000	5.9	18.5	1529000	50000	...	3.2	847000	201000	505000	97000	17000	21000
2	1447000	199000	7056000	4544000	2336000	393000	4.9	22.8	1615000	43000	...	2.8	743000	187000	453000	123000	152000	24000

3 rows × 21 columns

```python
# Check basic information about the dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3 entries, 0 to 2
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   vis-dom               3 non-null      int64  
 1   vis-int               3 non-null      int64  
 2   vis-night-dom         3 non-null      int64  
 3   vis-night-int         3 non-null      int64  
 4   exp-dom               3 non-null      int64  
 5   exp-int               3 non-null      int64  
 6   avg_length-stay-dom   3 non-null      float64
 7   avg_length-stay-int   3 non-null      float64
 8   avg-spend-trip-dom    3 non-null      int64  
 9   avg-spending-trip-int 3 non-null      int64  
 10  vis-mkt-share-dom     3 non-null      float64
 11  vis-mkt-share-int     3 non-null      float64
 12  purpose-holiday-dom   3 non-null      int64  
 13  purpose-fnf-dom       3 non-null      int64  
 14  purpose-business-dom  3 non-null      int64  
 15  purpose-others-dom    3 non-null      int64  
 16  purpose-holiday-int   3 non-null      int64  
 17  purpose-fnf-int       3 non-null      int64  
 18  purpose-business-int  3 non-null      int64  
 19  purpose-others-int    3 non-null      int64  
 20  year                  3 non-null      int64  
dtypes: float64(4), int64(17)
memory usage: 632.0 bytes
```

```python
# Check number of rows and columns in dataset
data.shape
```

```
(3, 26)
```

```python
# Check if there are any duplicates in dataset
data.duplicated()
```

	0
0	False
1	False
2	False

dtype: bool

```python
# Check if there is any missing data in dataset
missing = data.isnull().sum()
missing
```

	0
vis-dom	0
vis-int	0
total_visitor	0
vis-night-dom	0
vis-night-int	0
total_vis_night	0
exp-dom	0
exp-int	0
total_exp	0
avg_length-stay-dom	0
avg_length-stay-int	0
total_avg_length_stay	0
avg-spend-trip-dom	0
avg-spending-trip-int	0
total_avg_spend_per_trip	0
vis-mkt-share-dom	0
vis-mkt-share-int	0

[ ]

```python
# Tourist Prediction Data Analysis
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_squared_error

# Calculate total visitors (domestic + international)
data['total_visitors'] = data['vis-dom'] + data['vis-int']

# Calculate the year-over-year change in total visitors (increase or decrease)
data['visitor_change'] = data['total_visitors'].diff()

# Correlation analysis for feature selection
plt.figure(figsize=(10, 8))
sns.heatmap(data[features + ['visitor_change']].corr(), annot=True, cmap='coolwarm')
plt.title('Feature Correlation with visitor_change')
plt.show()

# Drop the rows with missing values in the 'visitor_change' column (the first row will have NaN)
data = data.dropna(subset=['visitor_change'])

# Define a more extensive feature set by including other relevant columns
features = ['total_visitors', 'vis-night-dom', 'vis-night-int', 'exp-dom', 'exp-int',
            'avg_length-stay-dom', 'avg_length-stay-int', 'avg-spend-trip-dom', 'avg-spending-trip-int']

X = data[features]  # Using multiple features now
y = data['visitor_change']

# Standardizing the data by scaling it
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42, shuffle=False)

# Initialize the Gradient Boosting Regressor model
model = GradientBoostingRegressor(n_estimators=100, random_state=42)

# Train the model on the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Evaluate the model using Mean Squared Error
mse = mean_squared_error(y_test, y_pred)
```

```python

# Display results
print("Mean Squared Error on test data:", mse)

# Display actual vs predicted results
result_df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
print(result_df.head())

```

iii. Heatmap for Feature Correlation with Number of Visitor Change:

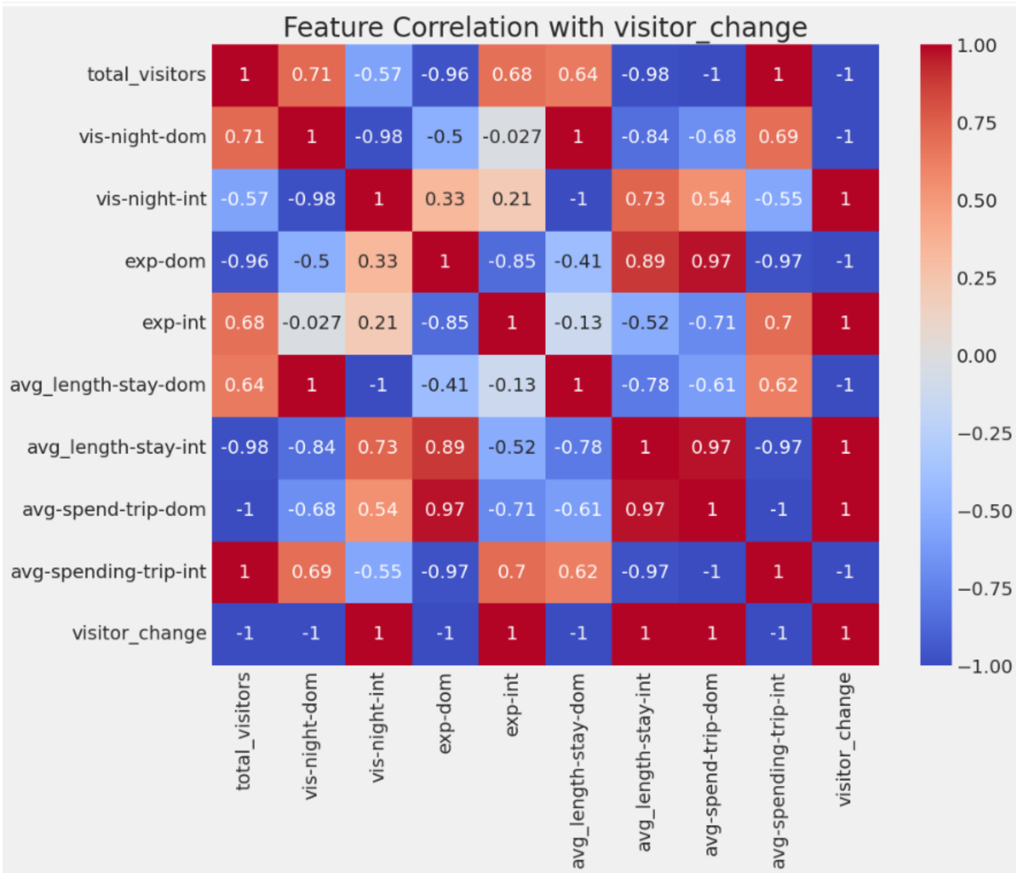


Figure 5: Heatmap of Prediction Model

iv. Output:

```
Mean Squared Error on test data: 36864000000.0
    Actual   Predicted
2 -55000.0   -247000.0
```

2. For NT Tourist Prediction Based on Each Location:

i. **Dataset:** We have created datasets from the source of NT state report of tourism on visitors from 2019-2024(March). All links are given in the reference section.

A	B	C	D
location	visitor_num	expenditure	year
GD	144000	1.242	2024
KD	301000	0.25	2024
LAS	282000	0.463	2024
KK	161000	0.227	2024
BA	46000	0.053	2024
ASM	348000	0.401	2024
GD	813000	1.16	2023
KD	291000	0.251	2023
LAS	297000	0.481	2023
KK	141000	0.181	2023
BA	48000	0.054	2023
ASM	333000	0.381	2023
GD	972000	1.631	2022
KD	210000	0.236	2022
LAS	237000	0.415	2022
KK	210000	0.236	2022
BA	77000	0.032	2022
ASM	303000	0.34	2022
GD	666000	1.015	2021
KD	257000	0.166	2021
LAS	159000	0.21	2021
KK	193000	0.162	2021
BA	57000	0.017	2021
ASM	268000	0.235	2021
GD	438000	0.38	2020
KD	221000	0.087	2020

Table 2: Dataset of NT Tourist based on Location

ii. **Python Code for Machine Learning Model:**

```
!pip install -q hvplot
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 155.4/155.4 kB 3.1 MB/s eta 0:00:00
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import hvplot.pandas
from scipy import stats

%matplotlib inline
sns.set_style("whitegrid")
plt.style.use("fivethirtyeight")
```

```
# Load the Tourist dataset
data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Govhack-data/Tourist-Location-based-Dataset.csv')
data
```

	location	visitor_num	expenditure	year
0	GD	144000	1.242	2024
1	KD	301000	0.250	2024
2	LAS	282000	0.463	2024
3	KK	161000	0.227	2024
4	BA	46000	0.053	2024
5	ASM	348000	0.401	2024
6	GD	813000	1.160	2023
7	KD	291000	0.251	2023
8	LAS	297000	0.481	2023
9	KK	141000	0.181	2023

```
# Check basic information about the dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 4 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   location     30 non-null     object 
 1   visitor_num  30 non-null     int64  
 2   expenditure  30 non-null     float64
 3   year         30 non-null     int64  
dtypes: float64(1), int64(2), object(1)
memory usage: 1.1+ KB
```

```
# Check number of rows and columns in dataset
data.shape
```

```
(3, 26)
```

```
# Get Statistical measures of the dataset
data.describe()
```

	visitor_num	expenditure	year
count	30.000000	30.000000	30.00000
mean	265300.000000	0.358967	2022.00000
std	214871.437103	0.391769	1.43839
min	46000.000000	0.017000	2020.00000
25%	141750.000000	0.130500	2021.00000
50%	215500.000000	0.235500	2022.00000
75%	300000.000000	0.396000	2023.00000
max	972000.000000	1.631000	2024.00000

```
# Check if there are any duplicates in dataset
data.duplicated()
```

```
0
```

```
[ ]
# Check if there is any missing data in dataset
missing = data.isnull().sum()
missing
```

	0
location	0
visitor_num	0
expenditure	0
year	0

dtype: int64

```
[ ]
# Location Based Tourist Prediction Data Analysis
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Initialize a dictionary to store the actual predicted numbers along with trends
future_predictions_with_numbers = {}

# Create a mapping for the location codes to their full names
location_mapping = {
    'GD': 'Greater Darwin',
    'KD': 'Katherine Daly',
    'KK': 'Kakadu Arnhem Land',
    'BA': 'Barkly',
    'LAS': 'Lasseter',
    'ASM': 'Alice Springs Macdonnell'
}

# Predict the number of tourists in 2025 for each location and show the predicted numbers along with the trend
for location in data['location'].unique():
    # Get the full name of the location
    full_location_name = location_mapping.get(location, location)

    # Filter data for the current location
    location_data = data[data['location'] == location]

    # Prepare the input features (year) and target (visitor_num)
    X = location_data['year'].values.reshape(-1, 1)
    y = location_data['visitor_num'].values

    # Train a linear regression model
    model = LinearRegression()
    model.fit(X, y)

    # Predict visitor numbers for 2025 and 2024
    prediction_2025 = model.predict(np.array([[2025]]))[0]
    prediction_2024 = model.predict(np.array([[2024]]))[0]

    # Determine the trend (increase or decrease)
    trend = "Increase" if prediction_2025 > prediction_2024 else "Decrease"

    # Store both the predicted number for 2025 and the trend
    future_predictions_with_numbers[location] = {
        "Predicted_Visitor_Num_2025": prediction_2025,
        "Tourism_Trend_2025": trend
    }

```

```python
    # Plot the actual data
    plt.figure()
    plt.plot(location_data['year'], location_data['visitor_num'], marker='o', label='Actual Visitor Numbers')

    # Plot the predicted data for 2025 as a separate point
    plt.scatter([2025], prediction_2025, color='green', label='Predicted Visitor Number (2025)', zorder=5)

    # Add title and labels
    plt.title(f'Tourist Numbers and Predictions for {full_location_name}')
    plt.xlabel('Year')
    plt.ylabel('Number of Tourists')

    # Show legend and display the plot
    plt.legend()
    plt.show()

# Convert to DataFrame for better visualization
future_trend_with_numbers_df = pd.DataFrame.from_dict(future_predictions_with_numbers, orient='index').reset_index()
future_trend_with_numbers_df.columns = ['Location', 'Predicted_Visitor_Num_2025', 'Tourism_Trend_2025']

# Replace the location codes with their full names in the future trend DataFrame
future_trend_with_numbers_df['Location'] = future_trend_with_numbers_df['Location'].map(location_mapping)

# Display the predicted numbers and trends for 2025
print(future_trend_with_numbers_df)
```

iii. Output:

	Location	Predicted_Visitor_Num_2025	Tourism_Trend_2025
0	Greater Darwin	474300.0	Decrease
1	Katherine Daly	314200.0	Increase
2	Lasseter	366000.0	Increase
3	Kakadu Arnhem Land	168800.0	Increase
4	Barkly	46500.0	Decrease
5	Alice Springs Macdonnell	401100.0	Increase

References:

1) https://www.tourismnt.com.au/research-strategies/research/latest-visitor-data
2) https://www.tourismnt.com.au/research/tra-summary-sheet-archive
3) https://www.tourismtopend.com.au/itinerary-planner?layout=itineraryplanner
4) https://northernterritory.com/things-to-do/art-and-culture/aboriginal-cultural-experiences
5) https://data.world/datasets/tourism
6) https://data.world/codefordc/airbnb-washington-d-c-2015-10-03
7) https://data.world/city-of-ny/rma9-fm39
8) https://data.nt.gov.au/dataset/?q=tourism&organization=department-of-primary-industry-and-resources&sort=score+desc%2C+metadata_modified+desc
9) https://data.nt.gov.au/dataset/nt-tourism-accommodation-provider-list-january-december-2020
10) https://data.nt.gov.au/dataset/nt-tourism-tours-provider-list-july-december-2020