



# Dataset Summary & Methodology

## Tourism Intelligence Model – The Bahamas

Data Years: 2024 and 2025

Prepared by: Dharma Data Solutions

---



### Dataset 1: `bahamas_pink_form_data_2024_weighted_realistic.csv`



#### Overview:

This dataset simulates the official immigration “Pink Form” collected from stopover air arrivals into **Nassau / New Providence** for the full year **2024**.



#### Logic and Methodology:

Attribute	Approach
Visitor Volume	Based on the official 2023 visitor count (4,441,540), adjusted for a <b>2% decline</b> in 2024 → total = <b>4,351,200 visitors</b>
Sampling	5,000 synthetic records were generated using Faker, each with a calculated <b>weight</b> to scale to the national total
Seasonality	Monthly distribution based on historical patterns from Bahamas Ministry of Tourism and Central Bank reports
Accommodation Type Distribution	<ul style="list-style-type: none"><li>Baha Mar Grand Hyatt: 33%</li><li>Atlantis Paradise Island: 23%</li><li>Boutique Hotels: 20%</li><li>Private Residence (Airbnb-style): 14%</li><li>Family/Friends: 10%  </li></ul> <b>  Length of Stay</b>   Randomized by accommodation type, aligned to observed averages (e.g., Baha Mar = 5–7 nights)   <b>  Country of Citizenship</b>   <ul style="list-style-type: none"><li>83% United States (assigned by major source states: FL, NY, TX, etc.)</li><li>7% Canada</li><li>2% UK</li><li>8% other countries  </li></ul> <b>  Purpose of Visit</b>   Weighted random assignment: Tourism, Wedding/Honeymoon,

Business, Education, Family |  
| **Weighting** | Each record includes a `Weight` field such that  $\text{SUM}(\text{Weight}) = 4,351,200$  |

---

## 2 Dataset 2: `resort_guest_data_2024_synced.csv`

### ✓ Overview:

A **perfect subset** of the 2024 Pink Form dataset, representing only guests who stayed at:

- Atlantis Paradise Island
- Baha Mar Grand Hyatt
- Recognized Boutique Hotels

### 🔧 Logic and Methodology:

Attribute	Approach
Data Source	Records were <b>filtered directly</b> from the pink form dataset by <code>Accommodation Type</code> and hotel name
Fields Preserved	Full Name, Arrival Date, Length of Stay, Country, Weight — fully aligned with pink form
Resort Type	Added as a new column: "Luxury Resort" vs. "Boutique Hotel"
Spend Simulation	<ul style="list-style-type: none"><li>• <b>ADR (Room Rate):</b> Simulated per resort (e.g., Baha Mar: \$400–\$650/night)</li><li>• <b>F&amp;B Spend/Day:</b> Based on typical guest behavior per tier</li><li>• <b>Ancillary Spend/Day:</b> Includes spa, waterpark, casino, golf, etc.</li><li>• <b>Total Estimated Spend:</b> Calculated by multiplying daily rates × nights stayed  </li></ul>

---

## 3 Dataset 3: `bahamas_pink_form_data_2025_weighted_realistic.csv`

### ✓ Overview:

Synthetic visitor data for **January to August 2025**, simulating growth and continued seasonality.

### 🔧 Logic and Methodology:

Attribute	Approach
Visitor Volume	Based on 2023 arrivals with a <b>5% growth</b> assumption → total = <b>4,662,000 visitors (Jan–Dec equivalent)</b>
Sampling	5,000 synthetic rows, weighted to scale to total YTD arrivals
Seasonality	Monthly weights reflect peak in March/July, lighter fall months, consistent with 2023–2024 trends
Accommodation Split	Same as 2024, assuming similar share distribution
Length of Stay & Citizenship	Assigned per accommodation and historical norms
Weighting	Each record includes a <code>Weight</code> field → $\text{SUM}(\text{Weight}) = 4,662,000$ visitors (scaled)

---

## **Dataset 4:** `resort_guest_data_2025_synced.csv`

### **Overview:**




A direct subset of the 2025 Pink Form data, including only resort and boutique hotel guests.

### **Logic and Methodology:**

Attribute	Approach
Subset Source	Pulled directly from pink form using <code>Accommodation Type</code>
Alignment	Every record is <b>1:1 aligned</b> to the pink form by: <ul style="list-style-type: none"> <li>• Full Name</li> <li>• Arrival/Departure Date</li> <li>• Length of Stay</li> <li>• Country</li> <li>• Weight  </li> </ul>   <b>Spend Simulation</b>   Same logic as 2024 resort dataset but with slightly adjusted (inflated) ADR ranges to reflect price increases in 2025

---

## **Use Cases**

-  Power BI dashboards for **seasonality, guest segmentation, and spending**
-  Economic impact modeling by **resort, month, and visitor type**
-  Reports for **Ministry of Tourism, Bahamas Hotel Association**, or external stakeholders



## File Outputs

File Name	Description
bahamas_pink_form_data_2024_weighted_realistic.csv	All visitors by air in 2024
resort_guest_data_2024_synced.csv	Subset of above for resort and boutique hotel guests
bahamas_pink_form_data_2025_weighted_realistic.csv	All visitors by air (Jan–Aug 2025)
resort_guest_data_2025_synced.csv	Subset of 2025 pink form data for resort and boutique stays