INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

EDA OF TOURS AND PACKAGES IN INDIA

About me

1.
- POLU SHRAVAN KUMAR
- BSC COMPUTERS
- Passionate in the Data Science Field. Data science is a dynamic and evolving field.
- LinkedIn: www.linkedin.com/in/shravan-kumar-polu-b5977b217
- GitHub: PoluShravanKumar (github.com)
- Fresher.

2.
- RANGU GURUCHARAN
- B.TECH (EEE)
- To solve real-world problems with data-driven solutions.
- LinkedIn: www.linkedin.com/in/rangu-gurucharan-ab5b0716b
- Fresher.

3.
- BAIRAM PAVAN
- BSC COMPUTERS
- Data science is one of the fastest-growing and most in-demand fields.
- LinkedIn: www.linkedin.com/in/bairam-pavan-b38624304
- Fresher.

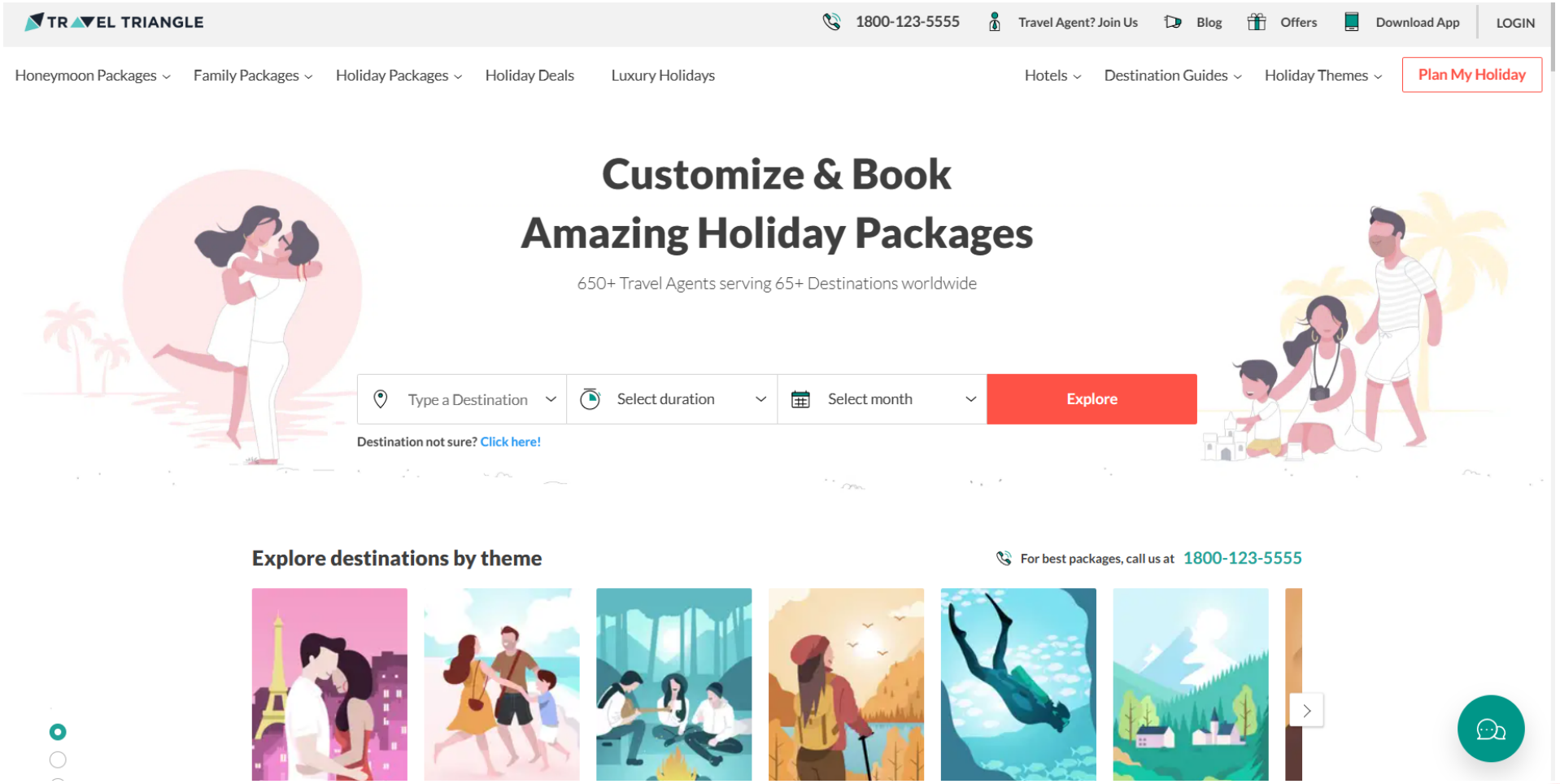
INNOMATICS RESEARCH LABS

Objective of the Project

- The Project Aim is to Understand the current landscape of tours and packages in India, including popular destinations, package types, pricing trends, and customer preferences.
- Analyze the offerings of competitors in the tourism industry to identify strengths, weaknesses, opportunities, and threats for your business.
- Analyze seasonal trends and peak booking periods for specific destinations or package types to optimize resource allocation and marketing efforts.

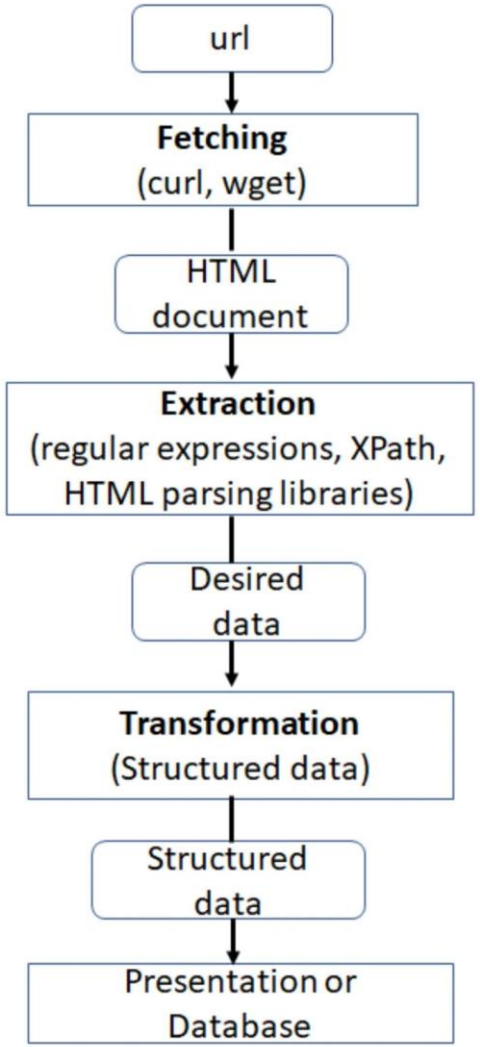
Web Scraping – Details (Websites, Processor you followed)

Site URL: [3650+ Best India Tour Packages - Upto 39% Off Domestic Packages (traveltriangle.com)]



INNOMATICS RESEARCH LABS

WEB SCRAPPING



- **Using RegEx, BeautifulSoup, Request libraries to fetch the HTML content of the website.**
- **This typically involves sending HTTP requests to the website, and parsing the HTML content.**
- **Selecting specific elements using CSS selectors or XPath expressions.**
- **After extracted the Data, saved it in a CSV file**.

INNOMATICS RESEARCH LABS

Most Reasonable Goa Honeymoon Tour Packages

Add To Compare ☐

5 Days & 4 Nights | Customizable

Starting from: 12% Off ⓘ

₹ 11,000/- ~~₹ 12,571/-~~

Per Person on twin sharing

Hotel included in package:

○ 3 Star ◉ 2 Star ○ 4 Star

Cities: Goa (5D)

Hot Pick | Water Activities | Romantic Cruise | Adventure

Check out TravelTriangle's list of 4 nights 5 days Goa honeymoon tour packages. Click here....

Upto 2 Stars | Meals | Sightseeing | +2 more

View Details

Customize & Get Quotes

Fantastic Andaman Honeymoon Package

Add To Compare ☐

5 Days & 4 Nights | Customizable

Starting from: 17% Off ⓘ

₹ 15,599/- ~~₹ 18,811/-~~

Per Person on twin sharing

Hotel included in package:

◉ 3 Star

Cities: Port Blair (3D) → Havelock (2D)

Corbyn's Beach Cove | Adventure | Radhanagar Beach | Nature

Spend your dream honeymoon by booking this 4 nights, 5 days Andaman honeymoon p....

Upto 3 Stars | Meals | Sightseeing | +2 more

View Details

Customize & Get Quotes

Summary of the Data

- DataFrame: Packages of Tours.
- Columns of the data frame: The data has 11 Columns and 875 Rows. The shape of the data frame is (875,11)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 875 entries, 0 to 874
Data columns (total 11 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Package_Type              875 non-null    object
 1   Duration_Days             875 non-null    int64 
 2   Duration_Nights           875 non-null    int64 
 3   Starting_Location         875 non-null    object
 4   Destination_Cities        875 non-null    object
 5   Destination_Cities_Count  875 non-null    int64 
 6   Included_Activities       875 non-null    object
 7   Hotel_Ratings             875 non-null    int64 
 8   Original_Price_INR        875 non-null    int64 
 9   Deal_Price_INR            875 non-null    int64 
 10  Discount_Percentage       875 non-null    int64 
dtypes: int64(7), object(4)
memory usage: 75.3+ KB
```

INNOMATICS RESEARCH LABS

Exploratory Data Analysis(EDA)

- Data after scrapping from the website.

	Package_Type	Duration_Days	Duration_Nights	Starting_Location	Destination_Cities	Destination_Cities_Count	Included_Activities	Hotel_Ratings	Original_Price_INR	Deal_Price_INR	Discount_Percentage
0	Most Reasonable Goa Honeymoon	5	4	Goa	Goa	1	Hot Pick Water Activities Romantic Cruise	2	12571	11000	12
1	Fantastic Andaman Honeymoon	5	4	Port Blair	PortBlair,Havelock	2	Corbyn's Beach Cove Adventure Radhanagar Be...	3	18811	15599	17
2	Romantic Mussoorie Tour Package	3	2	Mussoorie	Mussoorie	1	Sightseeing Nature Hill station Himalayas Roma...	3	12000	10500	13
3	Coorg Package	3	2	Coorg	Coorg	1	Coorg sightseeing Adventure Rajas Seat Nature...	3	17802	16200	9
4	Best Andaman Sightseeing Tour Package	5	4	Port Blair	PortBlair,Havelock	2	Cellular Jail Light & Sound Show Adventure	3	36484	33200	9
...	...	...	...	...	...	...	...	...	...	...	...
870	Golden Triangle Tour Package	2	1	New Delhi	NewDelhi	1	Adventure Nature Religious Budget	3	11702	11000	6
871	Alluring Assam Sightseeing Tour Package	6	5	Kaziranga	Kaziranga,Shillong,Guwahati	3	Sightseeing Historical Kaziranga National Pa...	2	31318	28499	9
872	Top-Selling Chennai Sightseeing Package	3	2	Chennai	Chennai,Mahabalipuram	2	Temple Religiousart gallery Historical Marina ...	3	16959	15433	9
873	Spectacular Shillong Holiday Package	3	2	Shillong	Shillong	1	Sightseeing Nature Temples Adventure Hills Hil...	3	13736	12500	9
874	Blissful Munnar Thekkady Alleppey Honeymoon	6	5	Munnar	Munnar,Thekkady,Kumarakom,Alleppey	4	Hill station Houseboat Stay Wildlife Beach W...	3	20909	18400	12

875 rows × 11 columns

INNOMATICS RESEARCH LABS

Exploratory Data Analysis

- **Data Cleaning**

- **It involves identifying and rectifying various issues such as missing values, duplicate records, outliers, and formatting inconsistencies.**
- **The main goals of data cleaning are to ensure that the data is accurate, complete, and consistent, making it suitable for analysis, modeling, and decision-making purposes.**

```
# Here all columns have 0 missing values.
Trips.isna().sum()
```

```
Package_Type                0
Duration_Days               0
Duration_Nights             0
Starting_Location           0
Destination_Cities          0
Destination_Cities_Count    0
Included_Activities         0
Hotel_Ratings               0
Original_Price_INR          0
Deal_Price_INR              0
Discount_Percentage         0
dtype: int64
```

- **There is no Duplicate values.**

```
# Replacing the invalid values with empty string.

list_Package_Type = []

for i in Trips["Package_Type"]:
    list_Package_Type.append(re.sub(r"â\x80\x9d","",i))

Trips["Package_Type"] = list_Package_Type

list_Package_Type1 = []

for i in Trips["Package_Type"]:
    list_Package_Type1.append(re.sub(r"â\x80\x9c","",i))

Trips["Package_Type"] = list_Package_Type1


list_Package_Type2 = []


for i in Trips["Package_Type"]:
    list_Package_Type2.append(re.sub(r"â\x80\x99","",i))

Trips["Package_Type"] = list_Package_Type2
```

```
word_replacements = {
    "  ":"",
    "arjeeling": "Darjeeling",
    "haramshala": "Dharamshala",
    "warka": "Dwarka",
    "epal": "Nepal",
    "elhi": "Delhi",
    "ainital": "Nainital",
    "hanaulti": "Dhanaulti",
    "orth": "North",
    "o ham": "Dho Dham",
    "ational": "National"
}

modified_tour_packages = []
for package in package_Type:

    for old_word, new_word in word_replacements.items():
        package = package.replace(old_word, new_word)
    modified_tour_packages.append(package)

modified_tour_packages
```

INNOMATICS RESEARCH LABS

```
Trips[(Trips["Deal_Price_INR"]<30000) & (Trips["Destination_Cities_Count"]>5) &(Trips["Duration_Days"]>10)]
```

	Package_Type	Duration_Days	Duration_Nights	Starting_Location	Destination_Cities	Destination_Cities_Count	Included_Activities	Hotel_Ratings	Original_Price_INR	Deal_Price_INR	Discount_Percentage
407	Soulful Char Dham Yatra Package	12	11	Haridwar	Haridwar,Barkot,Uttarkashi,Sitapur,Badrinath,SrinagarGarhwal	6	Sightseeing Trekking Temples Tour Hill station Hot	2	25108	23099	8
604	Char Dham Tour Package	12	11	Haridwar	Haridwar,Barkot,Uttarkashi,Sitapur,Badrinath,SrinagarGarhwal	6	Temple Tour Trekking Adventure Sightseeing Nature	2	25384	23099	9

- **Here the Data of Package prices below 30,000/-, Destinations are above 5, and Duration Days of Trip are above 10.**

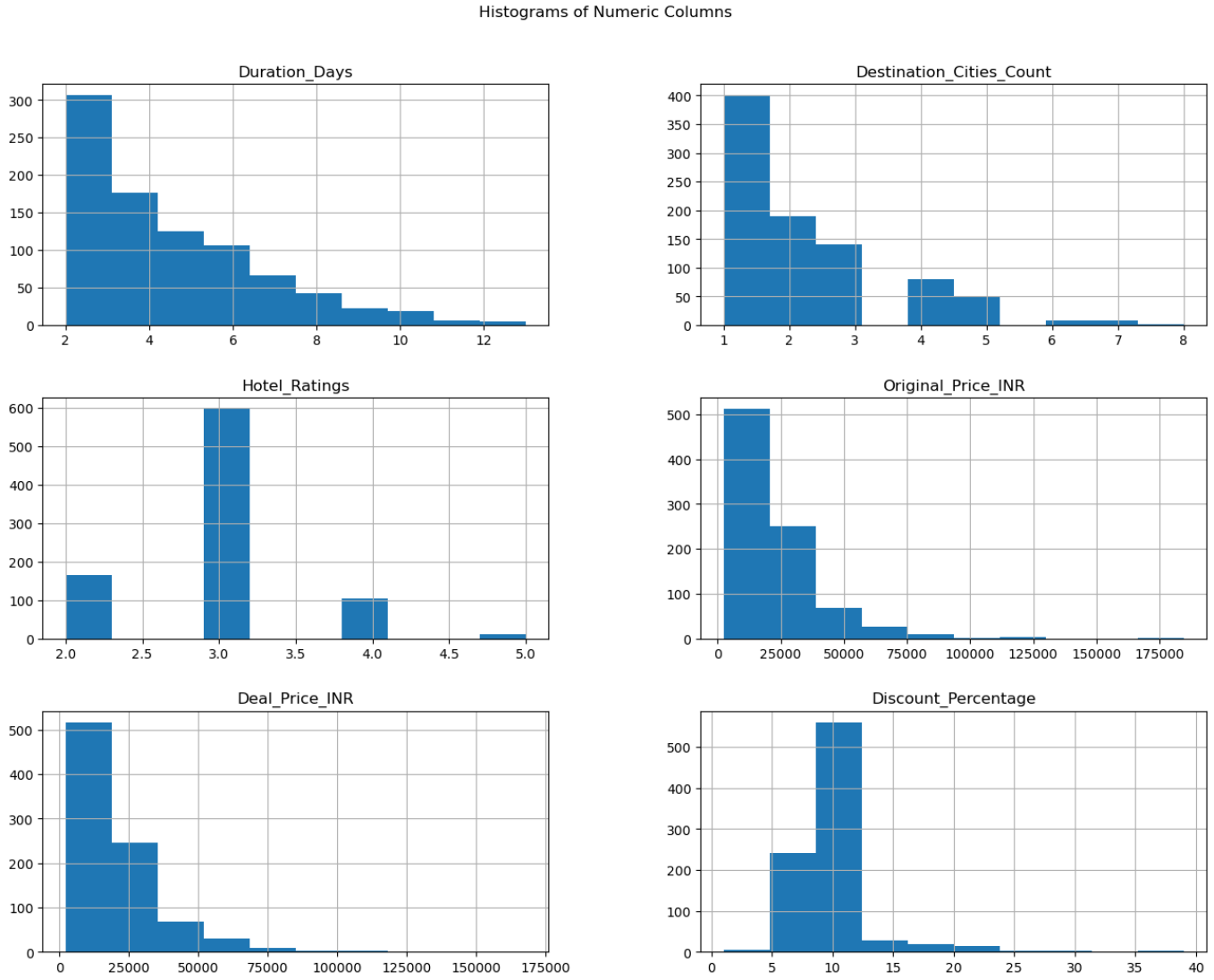
```
Trips["Starting_Location"].value_counts()[:20]
```

```
Starting_Location
Munnar          61
Gangtok         52
Goa             45
Shimla          39
Srinagar        38
Leh             36
Coorg           32
Ooty            32
Manali          29
Shillong        26
Port Blair      25
Mussoorie       20
Jaipur          19
New Delhi       18
Cochin          17
Alleppey        16
Udaipur         16
Lansdowne       15
Haridwar        15
Jim Corbett     15
Name: count, dtype: int64
```

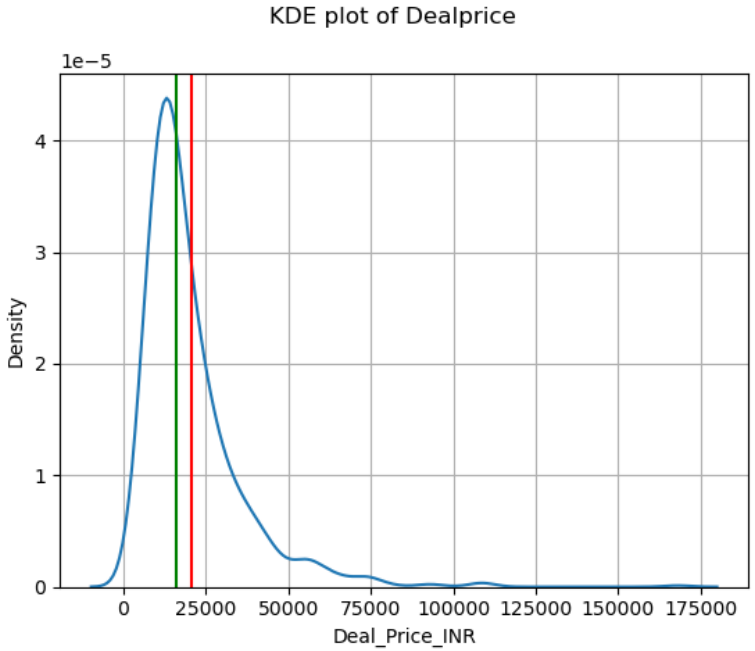
- **Here is the Data of the top 20 Starting locations.**
- **61 packages have a starting location at Munnar.**
- **People mostly like the tours of Munnar, Gangtok, and Goa.**
- **The Site mostly provides Munnar, Gangtok, and Goa compared with other Trips.**

Data Visualization

- Data visualization is the graphical representation of data to convey information and insights effectively.

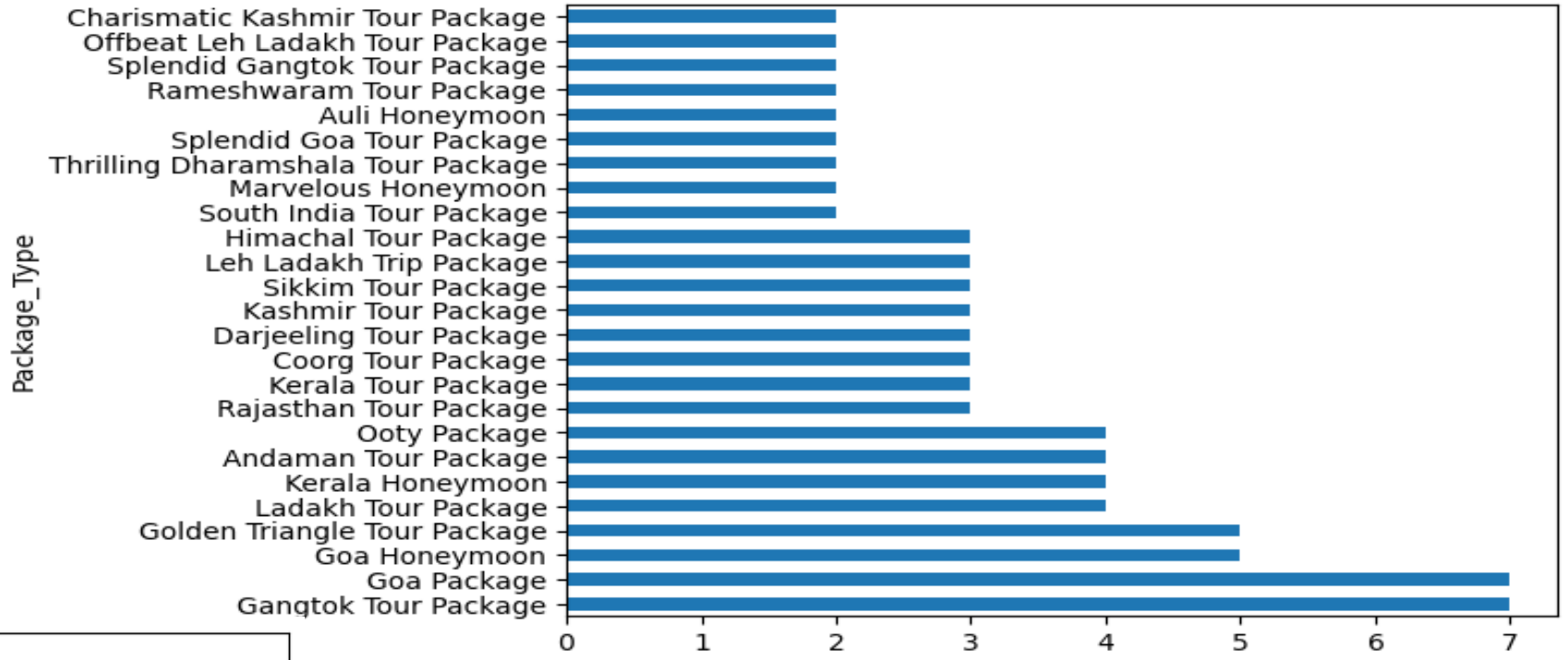


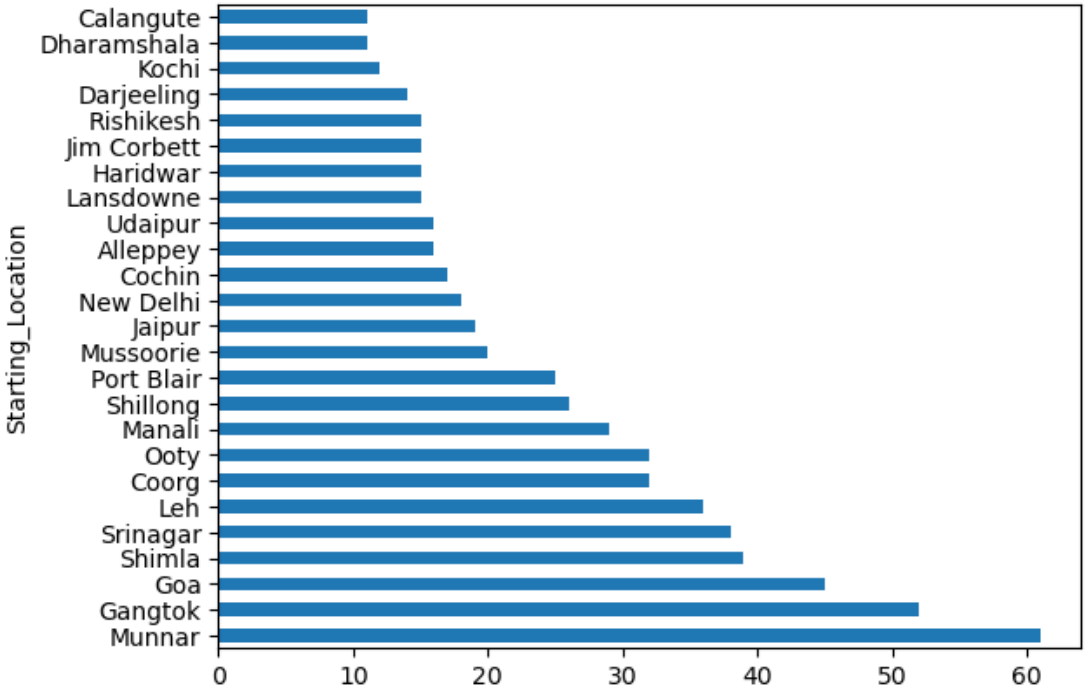
- Here are histogram graphs of all numerical columns.
- Histograms provide a visual representation of the distribution of numerical data.



INNOMATICS RESEARCH LABS

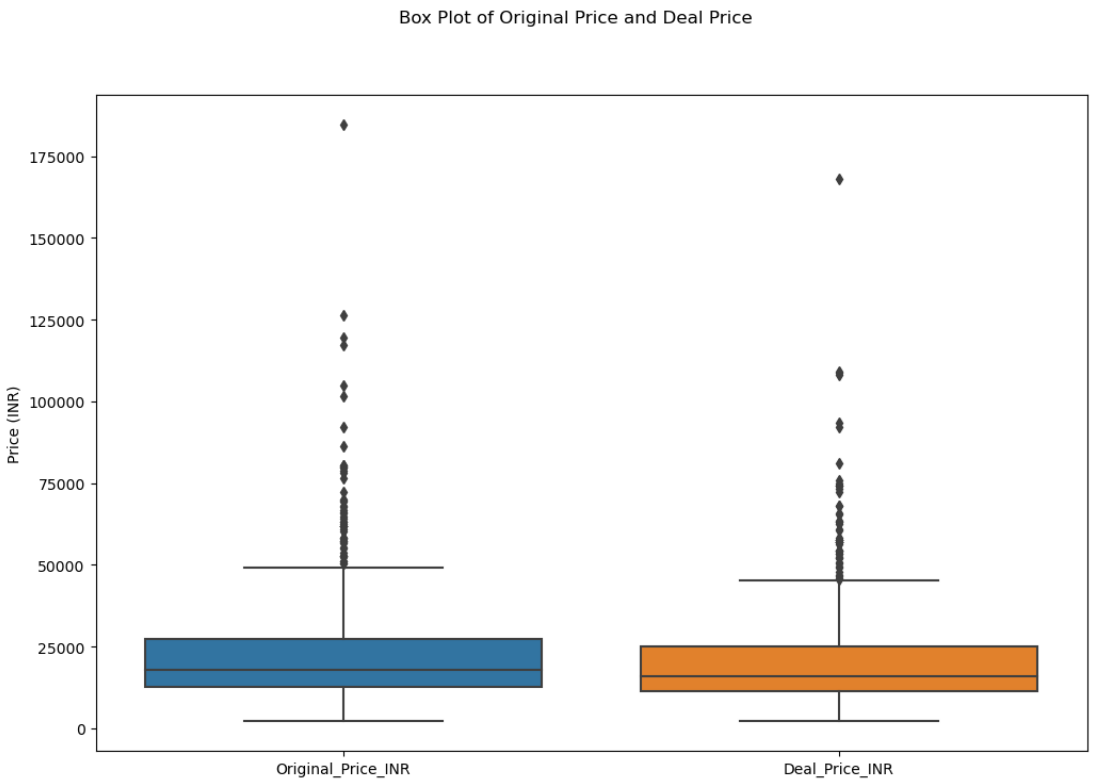
- **Bar plot of Top 25 packages.**
- Gangtok, Goa has more packages than other packages
- The site also provides more Goa and Gangtok packages.



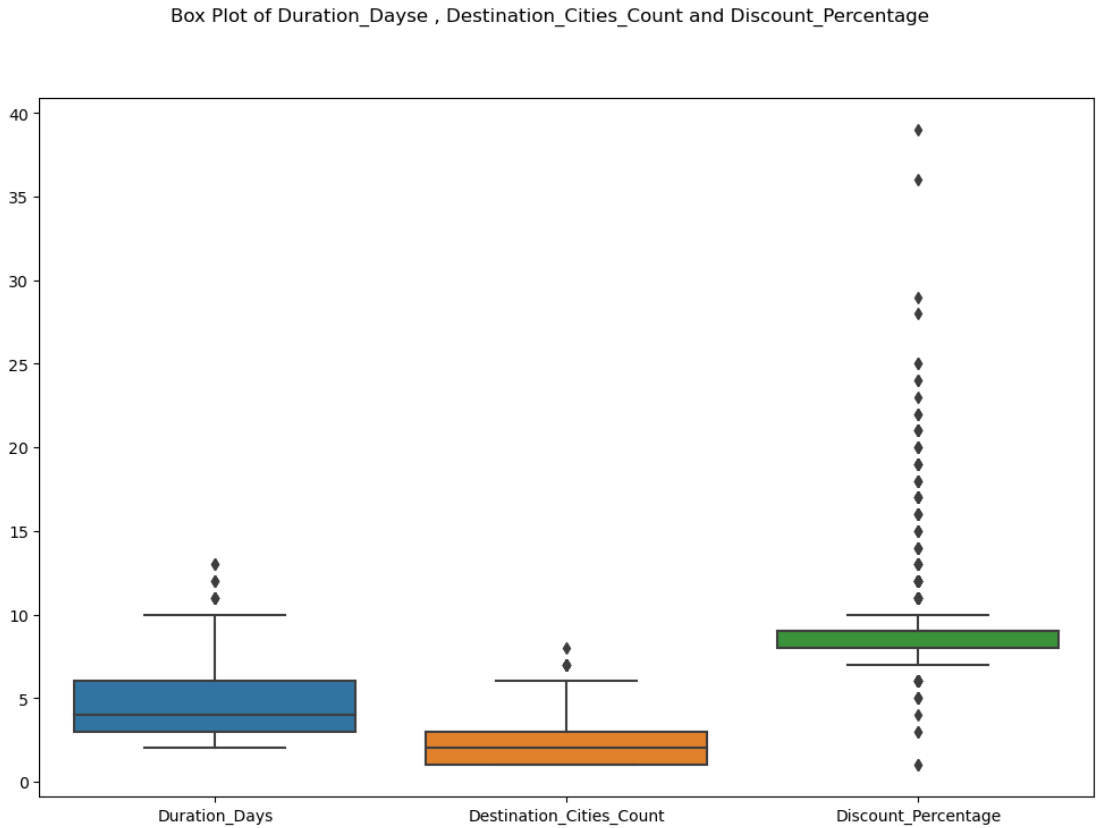


- **Bar plot of Top 25 Starting Locations.**
- The Site provides more packages that are starting from Munnar.

INNOMATICS RESEARCH LABS

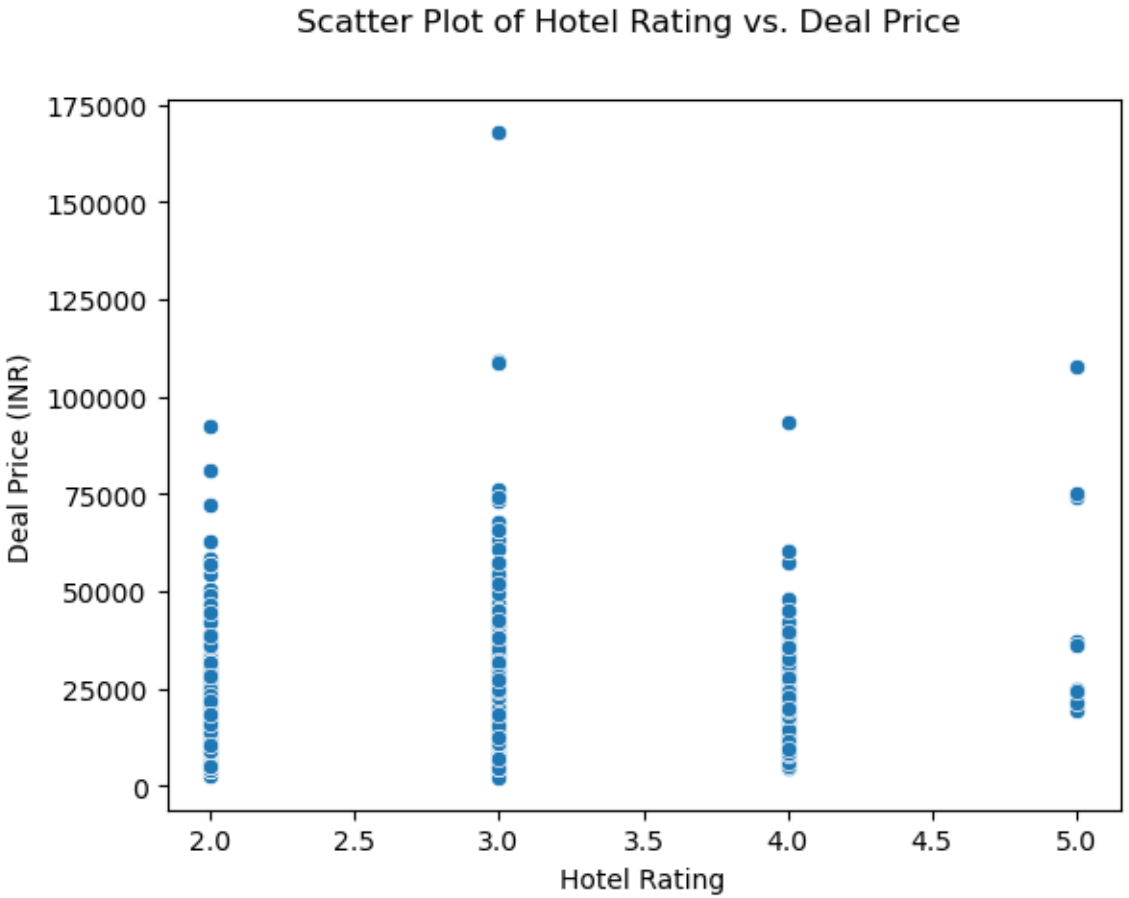


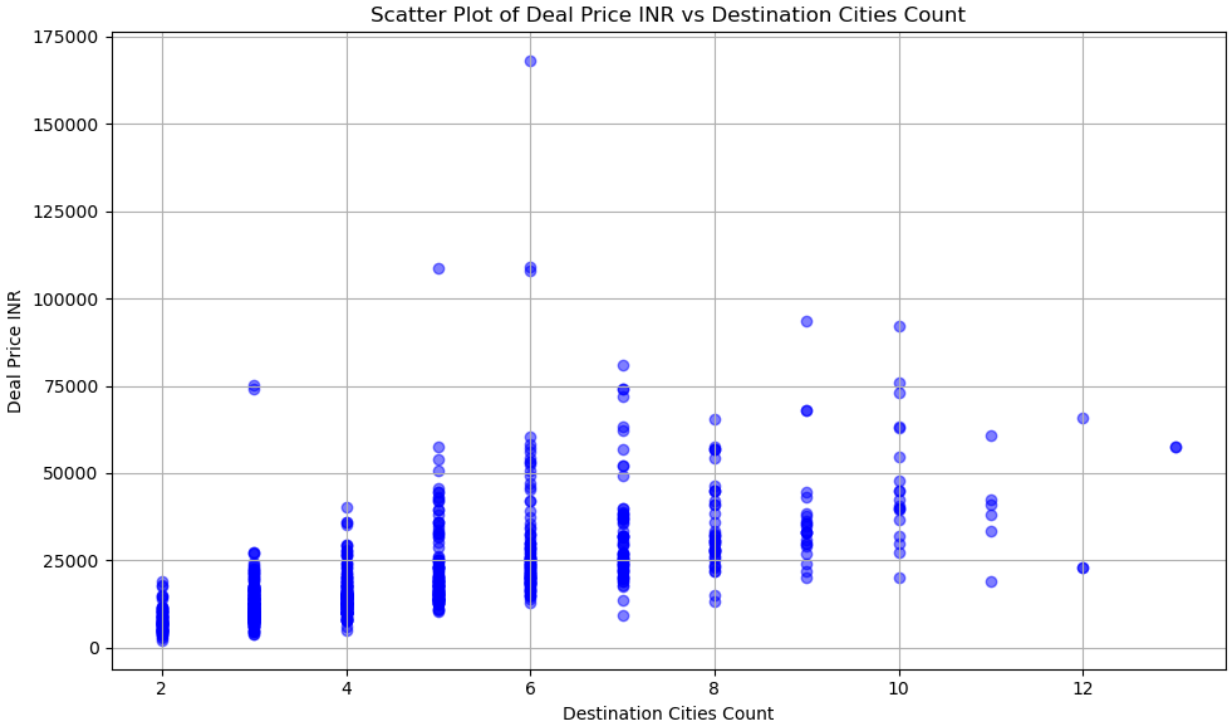
- **Box plots allow for a visual comparison of the distributions of the deal price and original price.**



- **Same here also we can see the Distributions of the Duration Days, Destination Cities Count, and Discount Percentage.**

INNOMATICS RESEARCH LABS

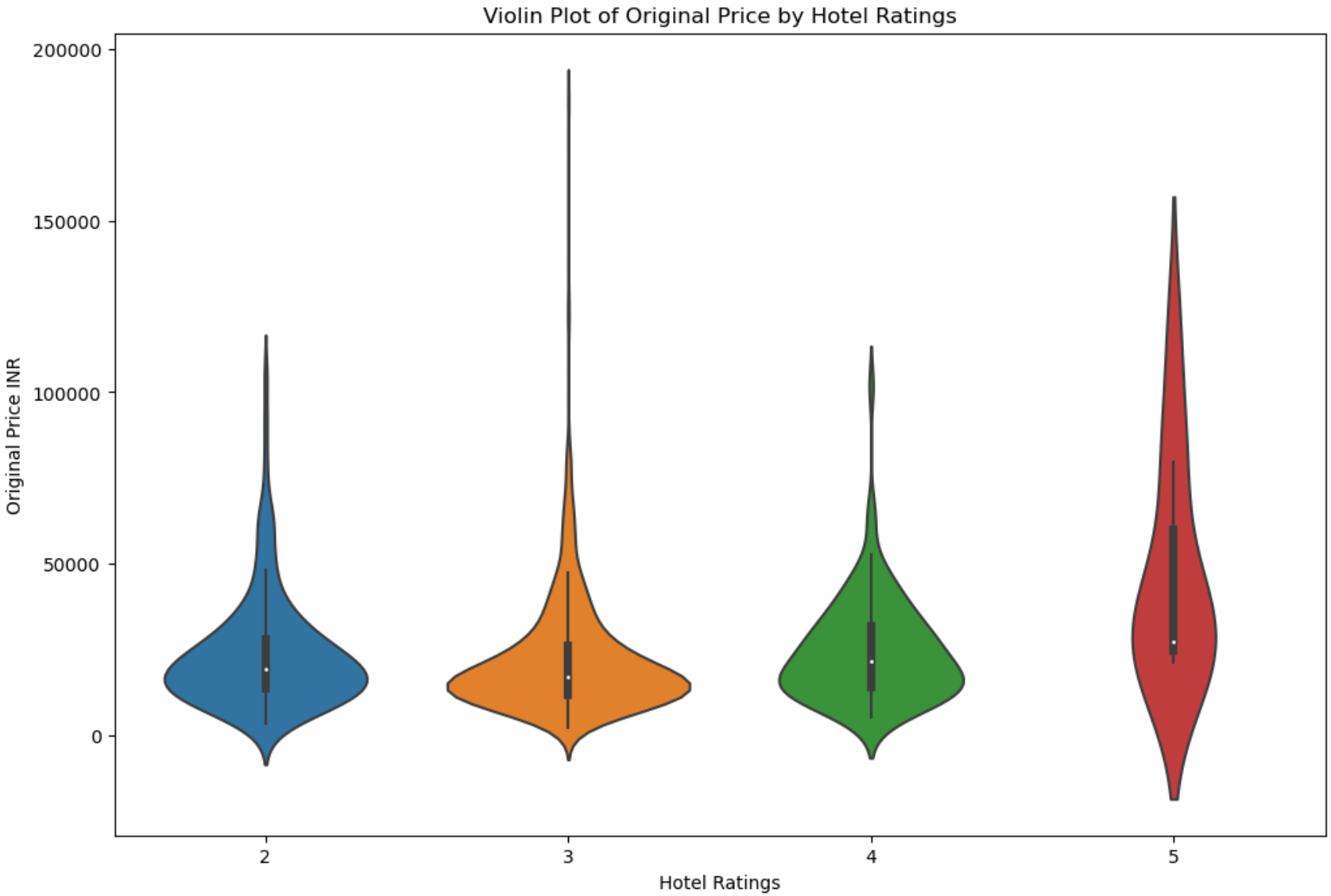




- **A positive relationship implies that as hotel rating increases, deal price tends to increase as well.**

- **Evaluate the strength of the relationship between destination city count and deal price.**

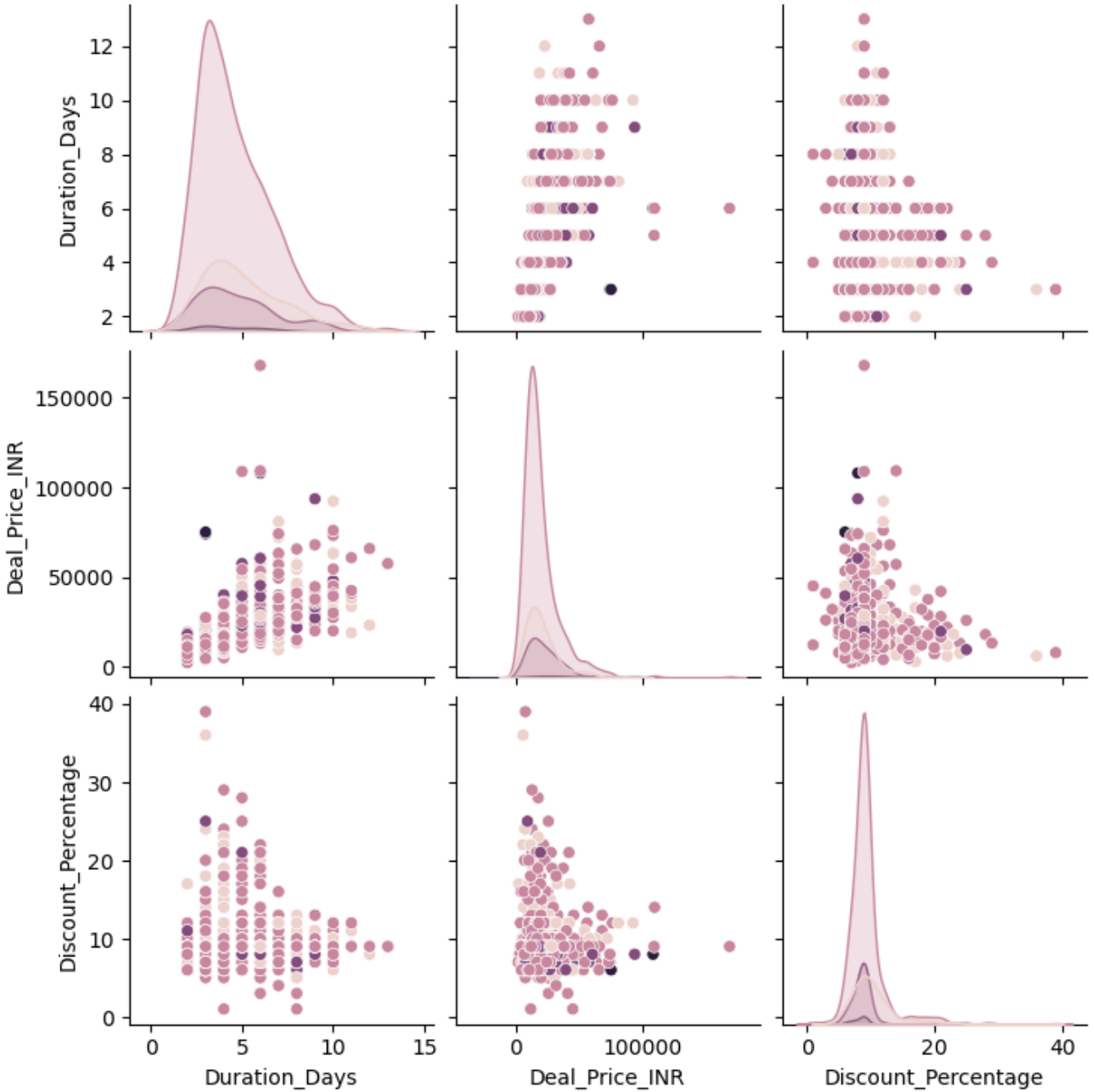
INNOMATICS RESEARCH LABS



- **Here the Relation Between Hotel Rating and Deal price with Violin plot.**

- **3 Stars Hotels are more in Trips Packages Compared to Other Rating hotels.**

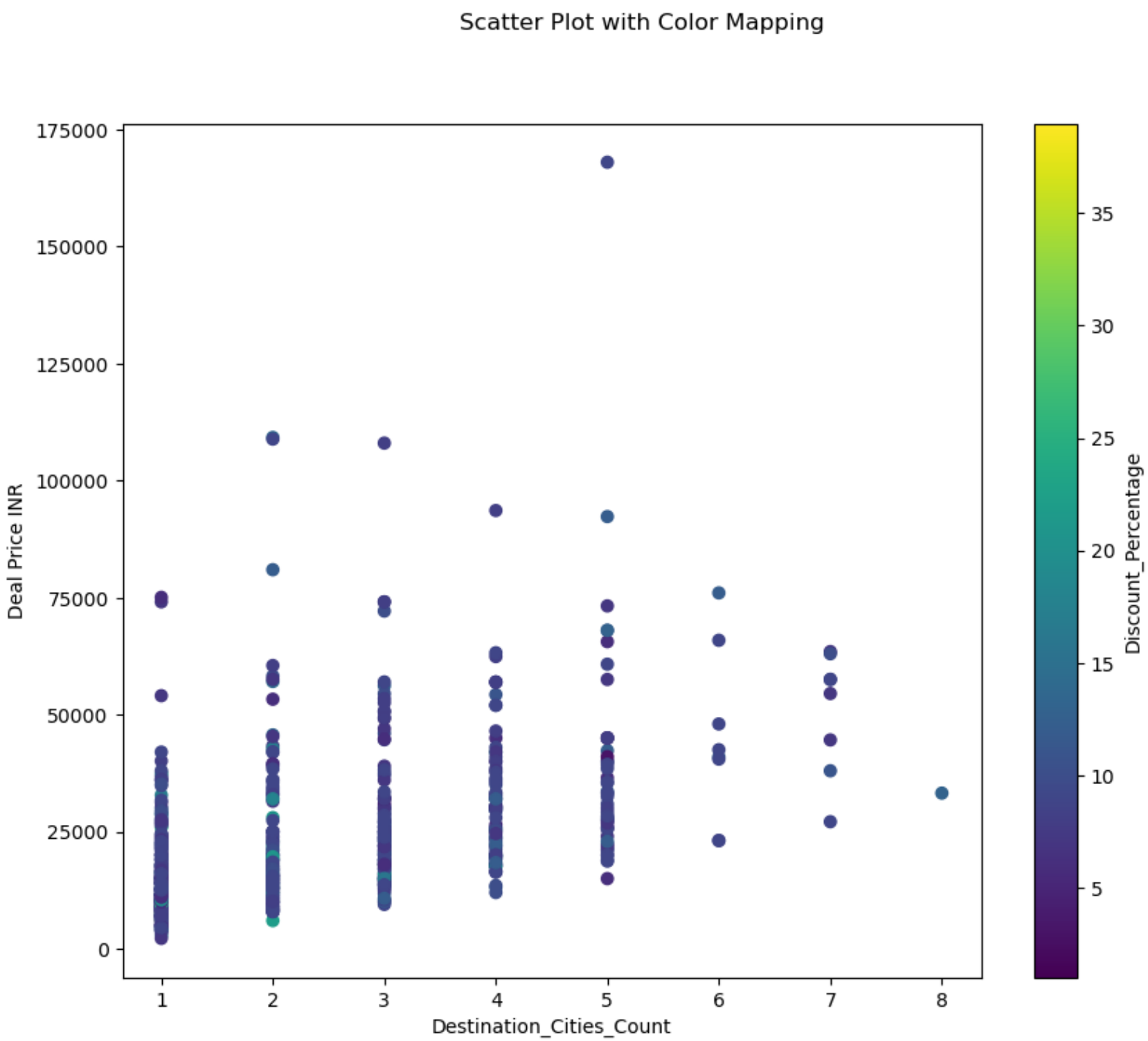
INNOMATICS RESEARCH LABS



- **Scatter Plot Matrix with Hue by Hotel Ratings**

- **Evaluate the relationships between the variables included in the pair plot (Duration_Days, Deal_Price_INR, Discount_Percentage, and Hotel_Ratings). Each scatter plot in the grid represents the relationship between two variables.**

INNOMATICS RESEARCH LABS



- **The color of each point represents the corresponding discount percentage (z-axis). Look for any patterns or trends in the color distribution across the scatter plot.**

INNOMATICS RESEARCH LABS

Key Business Question

- **Customer Demographics:** Investigate the demographics of customers who book tours and packages in India. Analyze factors such as age, gender, location, income level, and occupation to understand the profile of typical customers.
- **Popular Destinations**: Identify the most popular destinations among customers. Explore which cities, regions, or attractions attract the highest number of bookings and revenue.
- **Package Types**: Examine the types of tour packages offered and their popularity among customers. Investigate factors such as package duration, included activities, accommodation types, and price ranges.
- **Price Sensitivity**: Assess the impact of pricing on customer decision-making. Analyze how price levels, discounts, and promotional offers influence booking behavior and customer satisfaction.
- **Seasonality and Trends**: Investigate seasonal variations and trends in booking patterns. Analyze peak travel seasons, trends over time, and factors influencing seasonal fluctuations in demand.
- **Customer Preferences and Expectations**: Understand customer preferences, expectations, and motivations for booking tours and packages. Explore factors such as travel preferences, interests, trip purposes, and preferred travel styles.
- **Competitive Landscape**: Analyze the competitive landscape of the tours and packages industry in India. Identify key competitors, market share, pricing strategies, and unique selling propositions (USPs) of different providers.

The Conclusion

- **Demand Insights**: Based on the analysis, it can be concluded that there is a significant demand for Indian tours and packages across various demographic segments.
- **Popular Destinations**: Certain destinations emerge as clear favorites among tourists, indicating opportunities for tailored packages focusing on these locations.
- **Seasonal Optimization**: Understanding seasonal variations helps in optimizing package offerings and marketing strategies to attract tourists during off-peak seasons.
- **Budget Considerations**: Insights into budget allocation and spending patterns enable tour operators to design packages that cater to different budget ranges, enhancing accessibility to a broader customer base.
- **Competitive Positioning**: A thorough competitive analysis provides insights into market positioning and areas where differentiation can be achieved to gain a competitive edge.

✓ **In conclusion, leveraging the insights gained from EDA on Indian tours and packages is essential for developing targeted marketing strategies, optimizing package offerings, and enhancing customer satisfaction, ultimately driving growth and success in the tourism industry.**

INNOMATICS RESEARCH LABS

Q&A Slide

Q. What are the key demographic factors influencing the demand for Indian tours and packages?

Ans. Demographic factors such as age, gender, nationality, income level, and occupation influence the demand for Indian tours and packages. Younger demographics might seek adventure and nightlife, while older demographics may prefer cultural experiences and relaxation.

Q. Which are the top three most visited destinations in India according to the EDA?

Ans. Based on the EDA, the top three most visited destinations in India could be destinations like Gangtok for the Himalayas, Goa for its beaches, and Munnar for its rich cultural heritage.

Q. How do seasonal trends impact the influx of tourists to India?

Ans. Seasonal trends play a significant role in determining tourist traffic to India. For example, the winter season might see more tourists visiting northern hill stations, while the summer season might witness more visitors to coastal regions like Goa and Kerala.

Q. What insights does the EDA provide into the average budget allocated by tourists for Indian tours?

Ans. The EDA provides insights into the average budget allocated by tourists for Indian tours, including expenditure on accommodation, transportation, food, activities, and shopping. This helps in understanding the spending patterns of tourists and tailoring packages accordingly.

Q. How can sentiment analysis be used to enhance customer satisfaction in Indian tour packages?

Ans. Sentiment analysis can be used to analyze customer reviews and feedback to identify areas of improvement and strengths in existing tour packages. By addressing customer concerns and enhancing positive aspects, tour operators can improve overall customer satisfaction.

Experience

- **Data Exploration**: My journey begins with getting to know my dataset. I explore its structure, contents, and variables to understand what information it holds. This stage involves checking data types, identifying missing values, and gaining an initial understanding of the data's scope and complexity.

- **Formulating Questions**: As I delve deeper into the data, I start formulating questions and hypotheses to guide my analysis. These questions could be broad, exploring overall trends and patterns, or specific, targeting particular aspects of the data.

- **Visual Exploration**: Visualizations play a crucial role in EDA, allowing to uncover insights and patterns that may not be immediately apparent from the raw data. I create various plots, charts, and graphs to visualize relationships between variables, distributions of data, and trends over time.

- **Iteration and Refinement**: EDA is an iterative process. I often find myself revisiting earlier stages, refining my questions, and exploring different angles of the data as new insights emerge. This iterative approach allows me to deepen my understanding of the data and uncover hidden insights.

- **Statistical Analysis**: Alongside visual exploration, also performed statistical analysis to quantify relationships between variables, test hypotheses, and validate your findings. This could involve calculating summary statistics, conducting hypothesis tests, or fitting statistical models to the data.

THANK YOU

INNOMATICS RESEARCH LABS