

**Roll No:** 20BCE204

**Course Name and Course Code:** Big Data Analytics

**Practical:** 1

**Aim:** Study and explore various applications of big data in different domains. Prepare a study report for any select application which includes data sources, analytics and visualization algorithm.

## **1. Introduction**

'Big Data' refers to an extremely large volume of data and data sets that include structured and unstructured data coming in from multiple sources. These datasets are so voluminous that traditional data processing software is unable to capture, manage, or process them. Complex big data can be used to address business problems that were previously inaccessible.

Often, big data is characterized by the three Vs. – data containing great Variety, coming in increasing Volumes, with high Velocity. The data can come from publicly accessible sources like websites, social media, the cloud, mobile apps, sensors, and other devices. Businesses access such data to see consumer details like purchase history, what they searched for or what they watched, their likes, interests, and so on. Big data analytics uses analytic techniques to examine data, thus obtaining and find out information like hidden patterns, correlations, market trends, and consumer preferences. Therefore analytics help organizations make informed business decisions that lead to efficient operations, happy consumers, and increased profits.

## **2. Big Data in Travel Industry:**

Big Data has revolutionized the travel industry by leveraging the vast amount of data generated from travelers, booking platforms, hotels, airlines, and other related sources. Its applications have significantly transformed how travel companies operate, enhance customer experiences, and optimize their services.

- **Personalized Travel Experience:** Big Data enables travel companies to analyse vast amounts of customer data, including preferences, past behaviours, and feedback, to create personalized travel experiences. This personalization ranges from tailored travel recommendations and customized itineraries to targeted marketing campaigns.
- **Dynamic Pricing:** The travel industry extensively uses dynamic pricing models, where the cost of flights, hotel rooms, and other travel services changes based on demand and availability. Big Data analytics plays a crucial role in analysing real-time data and market trends to set optimal pricing strategies.
- **Fraud Detection and Security:** With the rise of online booking platforms, fraud and security concerns have also increased. Big Data analytics can identify suspicious transactions, patterns, and anomalies, helping travel companies mitigate fraud risks and ensure secure transactions.
- **Demand Prediction and Capacity Planning:** Big Data allows travel companies to predict demand patterns, identify peak travel periods, and optimize capacity planning for airlines, hotels, and other travel services. This ensures better resource allocation and maximizes revenue opportunities.
- **Sentiment Analysis and Customer Feedback:** Social media and online review platforms generate massive amounts of customer feedback. Big Data tools can analyse this unstructured data, providing valuable insights into customer sentiments and helping companies improve their services and offerings.

### **3. Data Sources of Travel Industry**

The travel industry is rich with diverse data sources that provide valuable insights for businesses, travelers, and policymakers. Some of the key data sources for the travel industry include:

- **Booking Platforms:** Online travel agencies (OTAs) like Expedia, Booking.com, MakeMyTrip, Goibibo, Ixigo and TripAdvisor gather massive amounts of data on hotel bookings, flight reservations, vacation rentals, and more. This data can be used to identify travel trends, popular destinations, and customer preferences.
- **Global Distribution Systems (GDS):** GDS platforms, such as Amadeus, Sabre, and Travelport, serve as intermediaries between travel suppliers (airlines, hotels, car rental companies) and travel agencies. They aggregate data on flight schedules, availability, and pricing, providing real-time information for bookings.
- **Customer Reviews and social media:** Platforms like TripAdvisor, Yelp, and social media networks are valuable sources of customer reviews, ratings, and feedback on hotels, restaurants, attractions, and destinations. Analyzing this data helps businesses improve their services and helps travelers make informed decisions.
- **Airline and Airport Data:** Airlines collect vast amounts of data on flight operations, passenger bookings, loyalty programs, and frequent flyer data. This data is crucial for airline revenue management, route planning, and customer service improvements. Airports generate data on passenger traffic, flight schedules, baggage handling, security wait times, and retail sales. This data is used to optimize airport operations and enhance the passenger experience.
- **Location-Based Services (LBS):** Mobile apps and platforms that provide location-based services offer data on travelers' movements, preferences, and interactions with local businesses. This data can help in personalizing travel experiences and targeted marketing.
- **Travel Surveys and Research:** Various organizations and market research firms conduct surveys and research on travel trends, spending patterns, and traveler preferences. For example, the World Tourism Organization (UNWTO) and travel industry associations regularly publish reports with valuable data.
- **Government and Tourism Boards:** Governments and tourism boards collect data on visitor arrivals, hotel occupancy rates, tourist spending, and tourism-related policies. This data helps in destination management and marketing strategies.
- **Weather and Climate Data:** Weather data is crucial for travel planning, especially for outdoor activities and popular tourist destinations. Weather forecasts and historical climate data are integrated into travel platforms.
- **Traffic and Transportation Data:** Data on traffic conditions, public transportation schedules, and transportation infrastructure are important for travelers navigating within cities or between destinations.

- **Geospatial Data:** Geographic information systems (GIS) like FlightRadar, FlightAware, etc provide spatial data about landmarks, points of interest, and geographical features, enriching travel applications with maps and navigation services.

It's essential to handle and process travel data responsibly, ensuring privacy and complying with relevant data protection regulations to gain actionable insights and deliver better travel experiences for customers.



#### 4. How is transportation a big data problem?

**Volume:** Refers to the vast amount of data generated in the transportation industry. The travel industry generates vast amounts of data from various sources, including travel bookings, reservations, customer profiles, loyalty programs, reviews, feedback, website interactions, mobile apps, social media, and more. The volume of data can range from terabytes to petabytes and beyond, especially for large global travel agencies, airlines, and hotel chains. Airbnb has over 5 million listings worldwide, offering a diverse range of accommodations to travelers. Booking.com features more than 28 million accommodation listings worldwide, catering to various travel preferences and budgets.

**Variety:** Travel industry data comes in various formats and types. Structured data, such as transactional records, booking details, and financial data, is commonly generated. Semi-structured data, like customer reviews, social media comments, and user-generated content, is also prevalent. Additionally, unstructured data, including images, videos, and audio files, are becoming increasingly important for sentiment analysis and customer feedback.

**Velocity:** Refers to the speed at which data is generated, collected, and processed in real-time or near-real-time. Travel companies need to process and analyze data quickly to respond to customer inquiries, handle bookings, manage inventory, and optimize pricing in real-time. Moreover, social media interactions, flight updates, and last-minute reservations contribute to the fast-paced data velocity. On average, Booking.com records over 1.5 million room nights booked every 24 hours, reflecting the rapid pace of bookings.

## 5. Data Analytics

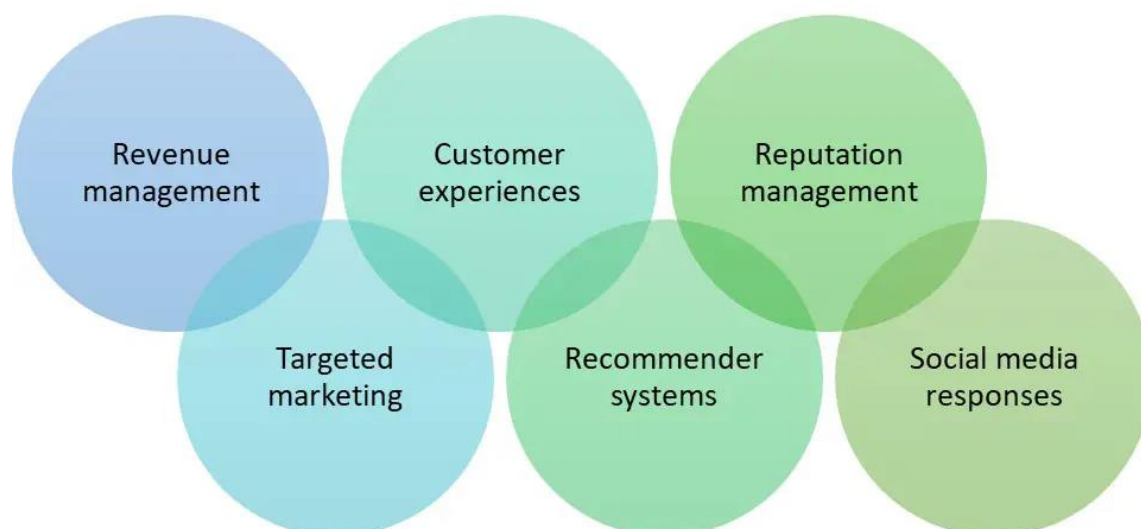
Data analytics unlocks many opportunities for travel companies. First and foremost, it allows people who are not data science experts to quickly review large-scale volumes of data. That is important because most of the touch points consumers have with travel businesses are now online, and each one produces some data. Big data and analytics can finally equip travel companies with everything they need to understand their target customers and capture more profit – or, in other words, gain a competitive advantage. Data analytics is essential because you can truly understand your business processes and how your company interacts with partners and customers.

Data science and big data have found many use cases in the tourism industry analytics. Travel organizations of all sizes are collecting and storing copious amounts of data. Some do it passively as their processes generate data.

For instance, their website can automatically collect data with Google Analytics to facilitate web analytics. They can also store interactions with customers and previous purchase history to personalize service and offer relevant products to consumers.

Others adopt a more active approach, engaging in data-gathering activities such as data mining. It can help them optimize pricing strategies or identify viable marketing channels for specific markets.

The primary role of big data in tourism industry analytics is to enable an accurate decision-making process. Proving accurate, updated, and structured data is paramount in this instance, as the quality analytics outcomes depend on all these factors.



## 6. Table:

Dataset	Description	Size	Insights
<a href="https://github.com/krishnaik06/5-Days-Live-EDA-and-Feature-Engineering/tree/main/Flight%20Prediction">https://github.com/krishnaik06/5-Days-Live-EDA-and-Feature-Engineering/tree/main/Flight%20Prediction</a>	The dataset contains travel information, including the airline, date of journey, source, destination, route, departure, arrival times, duration, total stops, additional details, and price, enabling comprehensive analysis of flight data for informed decision-making.	10684 datapoints	<ul style="list-style-type: none"><li>• <b>Flight Pricing Patterns:</b></li><li>• At time passengers are departing the most</li><li>• Which airline the fliers are preferring frequently</li><li>• Where the fliers are mostly travelling</li></ul>
<a href="https://www.kaggle.com/datasets/arianazmoud/airbnbopendata">https://www.kaggle.com/datasets/arianazmoud/airbnbopendata</a>	The New York Airbnb dataset comprises comprehensive listings of properties with detailed descriptions and average review scores, providing valuable insights for travelers seeking accommodation options in the city.	102600 datapoints	<ul style="list-style-type: none"><li>• Room type, construction year, and minimum nights provide valuable information on accommodation types and average guest stay durations.</li><li>• Review-related features (number of reviews, review rates, reviews per month) allow for sentiment analysis and evaluation of host performance, ensuring customer satisfaction.</li></ul>
<a href="https://www.kaggle.com/datasets/sanamptouropackageprediction?resource=download">https://www.kaggle.com/datasets/sanamptouropackageprediction?resource=download</a>	Customer-related attributes (Age, Gender, Occupation, MaritalStatus) and product details (TypeofContact, ProductPitched, PreferredPropertyStar) provide insights into customer behavior and preferences in the travel and hospitality industry. Additional factors like CityTier, MonthlyIncome, and OwnCar can aid in effective customer segmentation and targeted marketing strategies.	5000 datapoints	<ul style="list-style-type: none"><li>• Effective customer segmentation using factors like CityTier, MonthlyIncome, and OwnCar can help target marketing campaigns more efficiently.</li><li>• PitchSatisfactionScore provides insights into customer satisfaction, enabling businesses to optimize sales strategies and improve customer interactions.</li></ul>

## 7. References

- <https://www.optimizely.com/insights/blog/the-complete-guide-to-big-data-analysis-in-travel/>
- [https://www.techtarget.com/whatis/definition/3Vs#:~:text=The%203%20V's%20\(volume%2C%20velocity,number%20of%20types%20of%20data.](https://www.techtarget.com/whatis/definition/3Vs#:~:text=The%203%20V's%20(volume%2C%20velocity,number%20of%20types%20of%20data.)
- [https://www.hotelmize.com/blog/data-analytics-in-tourism-industry-what-is-it-benefits-how-its-used-real-life-examples/#What are the benefits of having data analytics in the tourism industry](https://www.hotelmize.com/blog/data-analytics-in-tourism-industry-what-is-it-benefits-how-its-used-real-life-examples/#What%20are%20the%20benefits%20of%20having%20data%20analytics%20in%20the%20tourism%20industry)