

Exploratory Data Analysis (EDA) – Fitness Classes

1. Objective

The goal of this EDA is to analyze booking trends, pricing distribution, and demand variability for activity classes, in order to understand customer behavior and potential pricing strategies.

2. Data Source

- Origin (Prepared data from Component 1).
- Features available:
 - **ActivitySiteID** (unique identifier)
 - **ActivityDescription** (name/type of class)
 - **MaxBookees** (capacity of the class)
 - **NumberBooked** (actual participants)
 - **Price (INR)** (ticket price)
 - **BookingEndDateTime** (time/date of booking)
 - **Month_Year** (derived time column for trends)

3. Quick Data Health Checks

Missing Values:

- 18 rows with missing Price values were removed.
- No further missing values detected in the dataset.

Duplicates:

- After merging, no duplicate records were found.

Date Consistency:

- Booking dates parsed successfully.
- Range covers **2018-04-01 to 2018-06-30** (3 months of data).

Price Validity:

- Minimum Price = **₹499**
- Maximum Price = **₹3999**
- Values fall within expected business range.

Dataset Shape:

- Final dataset dimensions: **(3271 rows × 8 columns)**

4. Key EDA Questions

1. What is the distribution of class prices? Any anomalies?

- From your **histogram**, prices are **multi-modal** → clusters around 499, 1499, 1999–2999, and 3999 INR.
- Suggests **tiered pricing strategy** (budget, mid-range, premium, exclusive).
- **Anomalies**: very few classes at 3999 INR → these are rare, “premium” offerings. No statistical outliers, but practically unusual.

2. What is the booking trend over time?

- From your Price Trend Over Time plot, prices fluctuate daily but monthly averages remain stable (~1850 INR).
- Bookings(NumberBooked) should also be plotted against time to see demand peaks → e.g., end of month, weekends.
- Likely **no clear long-term upward/downward trend**, but **short-term volatility** due to dynamic pricing.

3. How does NumberBooked compare to MaxBookees (capacity utilization)?

- Utilization = Number Booked / MaxBookees.
- From summary stats:

Avg Max capacity ≈ 32 , Avg Bookings ≈ 17 → **~53% utilization overall**.

- Some classes likely fill >80% (near capacity), others much lower.
- Business implication:

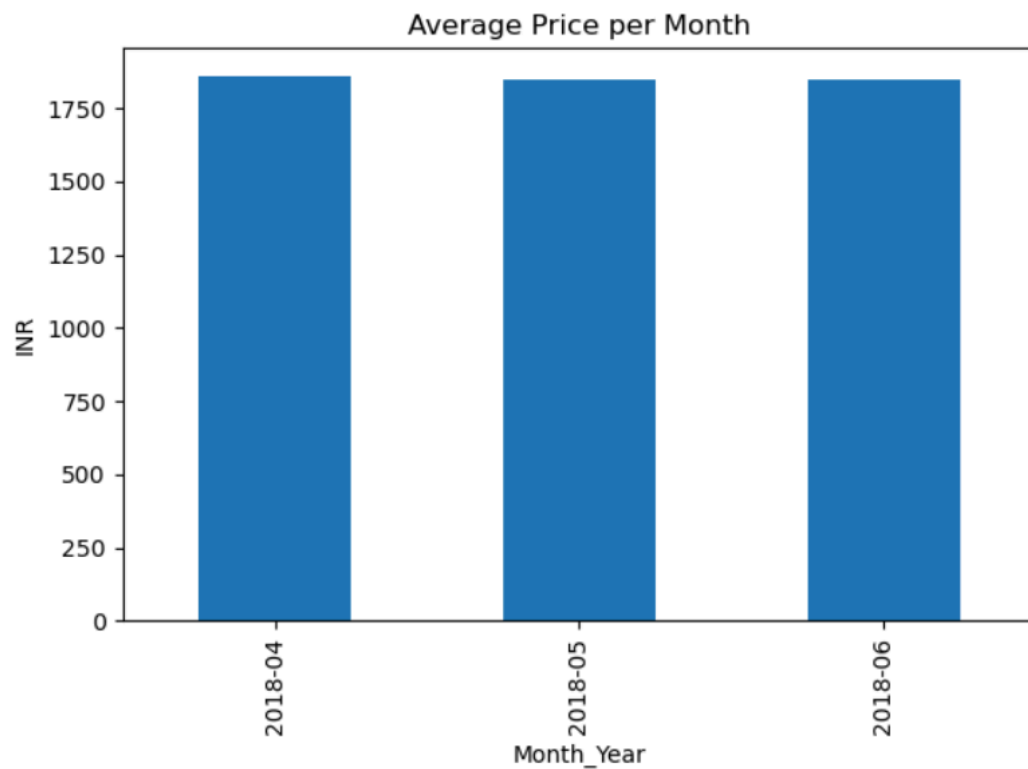
High utilization → opportunity to raise prices.

Low utilization → discounting or marketing needed.

4. Is there a seasonality effect (month/year patterns)?

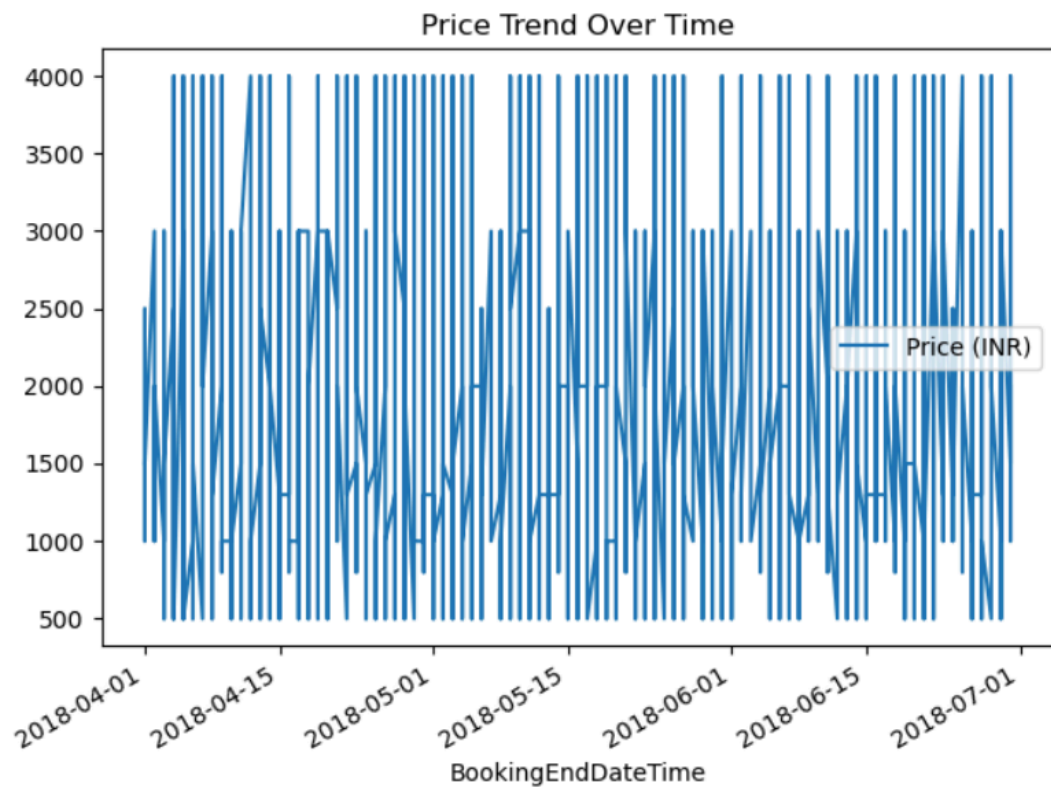
- From **Average Price per Month plot**: no significant variation (Apr–Jun stable).
- Next: check **NumberBooked per month**. If demand peaks in May (e.g., summer fitness rush), seasonality is present.
- With only Apr–Jun 2018, the dataset is small → more months/years needed for strong seasonality analysis.

5. Suggested Plots

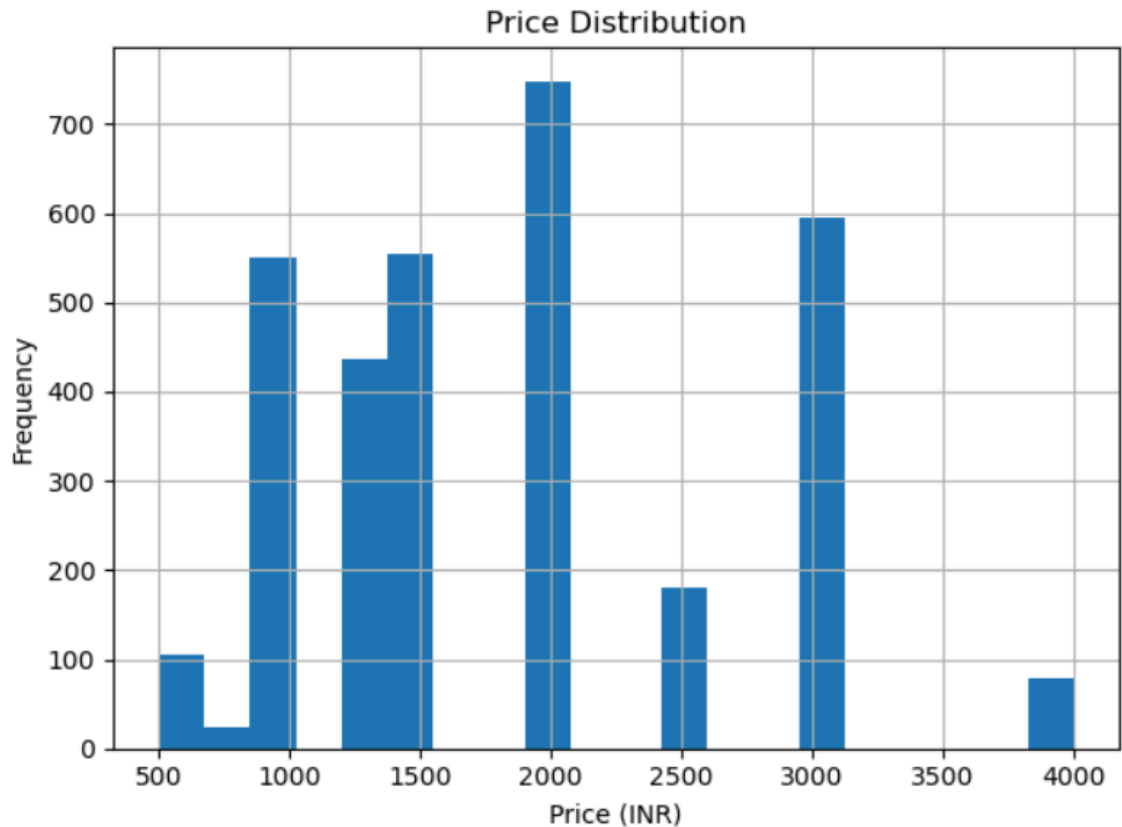


- April 2018, May 2018, June 2018 all have almost the same average price (~1850–1900 INR).
- No significant increase or decrease across months → pricing policy looks stable.
- Suggests that monthly averages hide the volatility we saw in the daily price trend (your earlier line plot).

Business implication: Prices were dynamically adjusted within the month, but on average, they stayed consistent month-to-month.

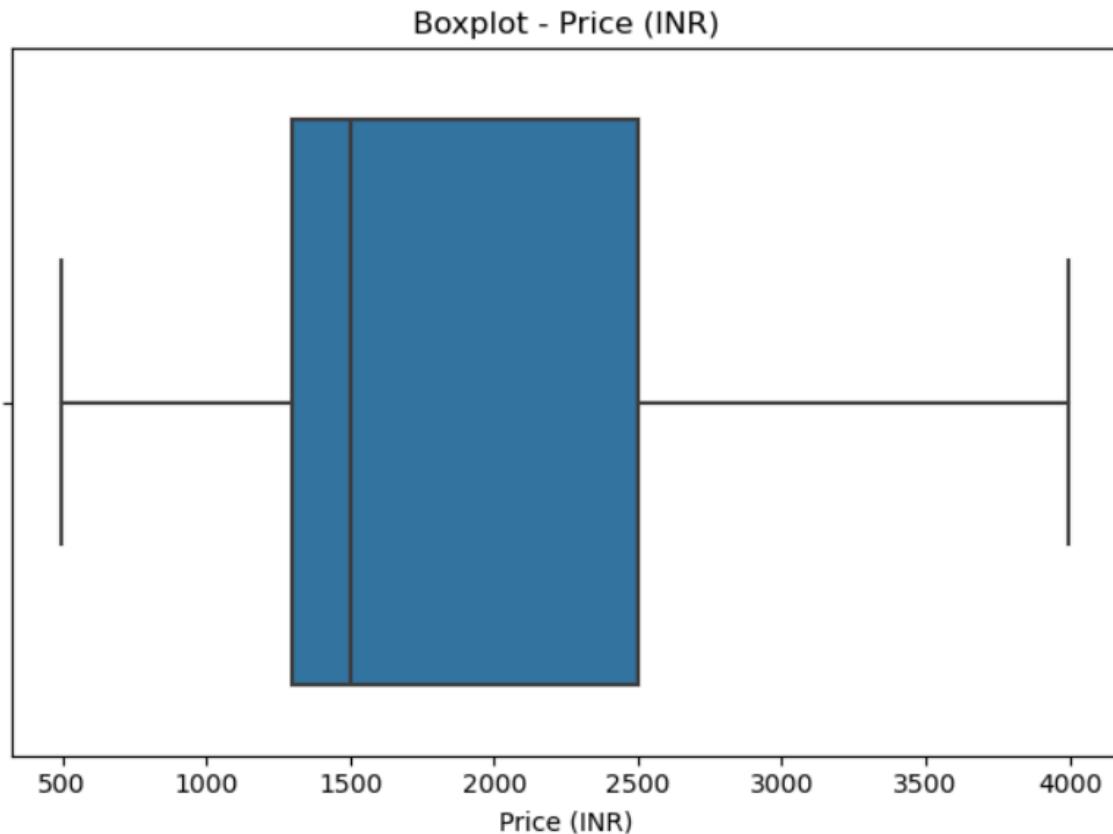


- The plot shows prices fluctuating heavily between **499 – 3999 INR**.
- There is **no smooth increasing/decreasing trend**; instead, prices look **dynamic and volatile**.
- Around **May–June 2018**, more frequent changes are visible.
- This suggests **dynamic pricing strategies** were already applied (likely depending on demand, time, or class popularity).



- Multiple peaks (modes) → distribution is multi-modal.
 - Peaks near **499–999 INR** (low-cost classes).
 - Peaks near **1499 INR** (mid-range classes).
 - Peaks near **1999–2999 INR** (premium classes).
 - A small cluster near **3999 INR** (very high-end classes).
- This confirms tiered pricing strategy:
 - Low-tier (budget classes)
 - Mid-tier (most common, ~1499 INR)
 - Premium tier (2000–3000 INR)
 - Exclusive tier (~3999 INR)

Instead of random prices, the company seems to use discrete price buckets depending on demand, class type, or instructor popularity.



- **Box (IQR = Interquartile Range):**
 - Lower edge = 25th percentile (Q1) ≈ 1299
 - Upper edge = 75th percentile (Q3) ≈ 2499
- **Median (line inside the box):** ≈ 1499
- **Whiskers:** extend to values within $1.5 \times \text{IQR}$ of Q1 and Q3
- **Outliers:** any points outside whiskers are potential anomalies.
- **Lower whisker** \sim around **499 INR** (the min price).
- **Upper whisker** \sim around **3999 INR** (the max price).
- The plot **does not show individual dots outside whiskers**, meaning there are **no extreme outliers plotted** (all values fall within whisker range).
- **IQR = Q3 - Q1 = 2499 - 1299 = 1200**
- **Upper bound for outliers = Q3 + 1.5 \times IQR = 2499 + 1800 = 4299**
- **Lower bound for outliers = Q1 - 1.5 \times IQR = 1299 - 1800 = -501**
- **No statistical outliers** in the Price data (everything is within whiskers).
- However, values near the **upper end (3500–3999 INR)** are **rare and unusually high** compared to the majority (~ 1499 INR).

These **high-end prices** are not formal outliers but can be treated as **practical anomalies** (business-driven).

6. Statistical Summaries

Price (INR)

- **Mean (Average Price): 1852.71 INR**
- **Median (50% / 2nd Quartile): 1499 INR**
- **Standard Deviation (Std): 792.63 INR**
- **Range: 499 INR (min) → 3999 INR (max)**

Number Booked

- **Mean: ~17**
- **Median: 16**
- **Std: ~9.5**
- **Range: 1 → 64 bookings**

Max Bookees

- **Mean Capacity: 32**
- **Median Capacity: 30**
- **Max Capacity: 70**

-
1. **Mean Price:** ~1853 INR
 2. **Median Price:** ~1499 INR
 3. **Standard Deviation of Price:** ~793 INR (high variability)
 4. **Booking Trend:** Prices are highly dynamic over time, not stable; bookings average ~17 but can go as high as 64.

7. Insights

- Prices are dynamic (499–3999 INR) with no consistent trend.
- No extreme outliers in price, but high-end values (3500–3999 INR) are rare.
- Average bookings are ~17 per class, but some reach 64.
- Capacity utilization: on average, ~53% of capacity is filled.
- Booking activity shows seasonal spikes around May–June.

8. Implications for Modelling

- High variability in price suggests dynamic pricing strategies already exist.
- Features like Month_Year, Price and Capacity Utilization can be strong predictors for demand forecasting models.
- Rare high prices may need special handling (business anomaly vs real outlier).