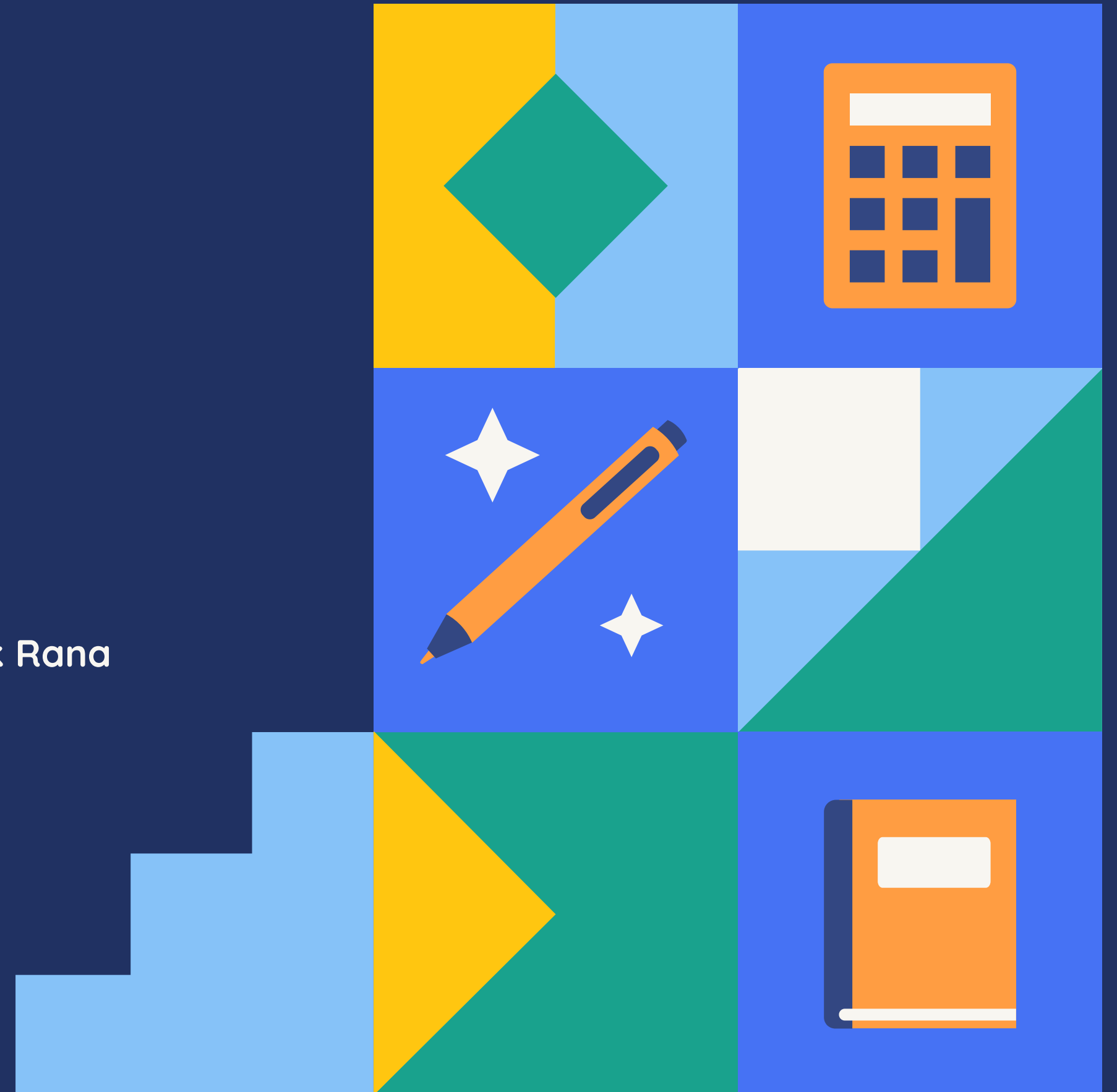
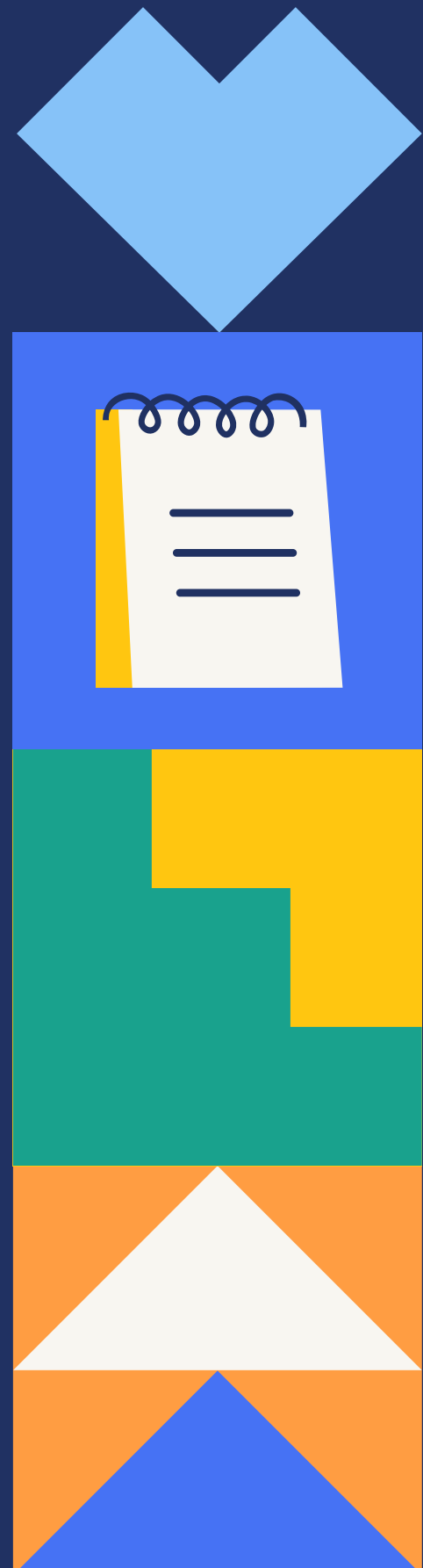


# CSD-103

By:- Suncoders

Akshat Gupta, Venkateshwar Sahu, Ansh Arora, Anurak Rana





# Outline



## Presenting

- Introduction
- Exploratory Data Analysis

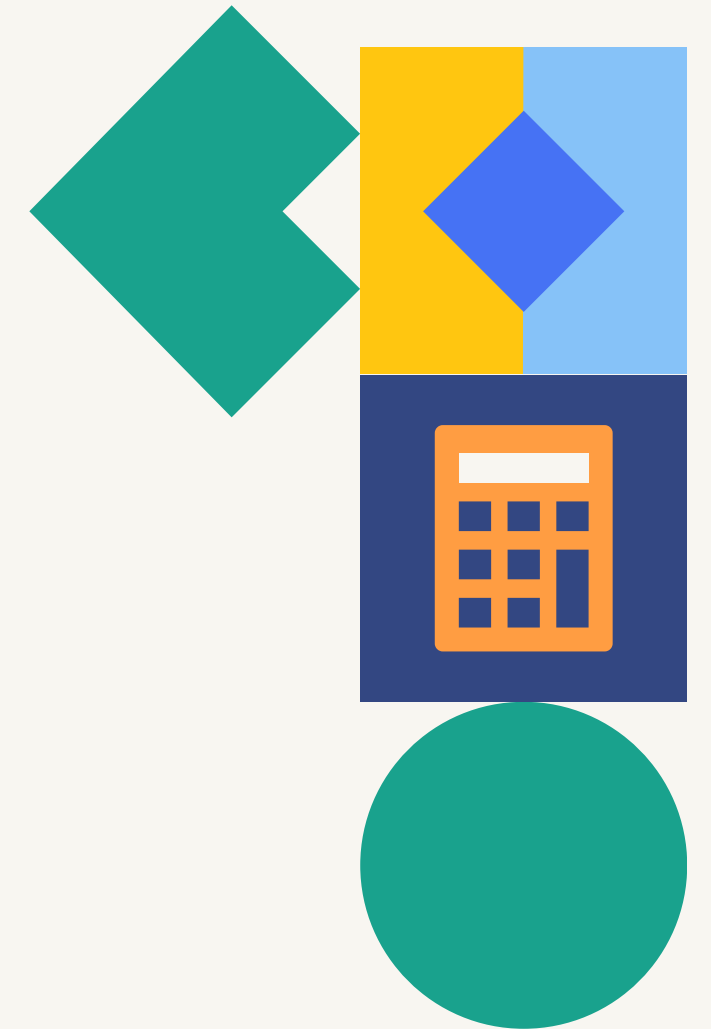


## Summarizing Results

- Scenarios
- Conclusion

# Introduction

- **The dataset under consideration encompasses a myriad of variables, including sales information such as units sold, unit prices, total revenue, costs, and profits, alongside categorical data like regions, item types, sales channels, and order priorities. To illustrate the practical application of statistical concepts, a series of hypothetical scenarios has been devised, each designed to demonstrate the crucial role of sampling techniques and statistical inferences in deriving valuable insights and making informed decisions.**
- **Through these scenarios, the objective is to showcase the significance of these statistical tools in extrapolating insights from limited but representative samples, thereby enabling data-driven decision-making and enhancing the understanding of larger datasets in real-world applications.**



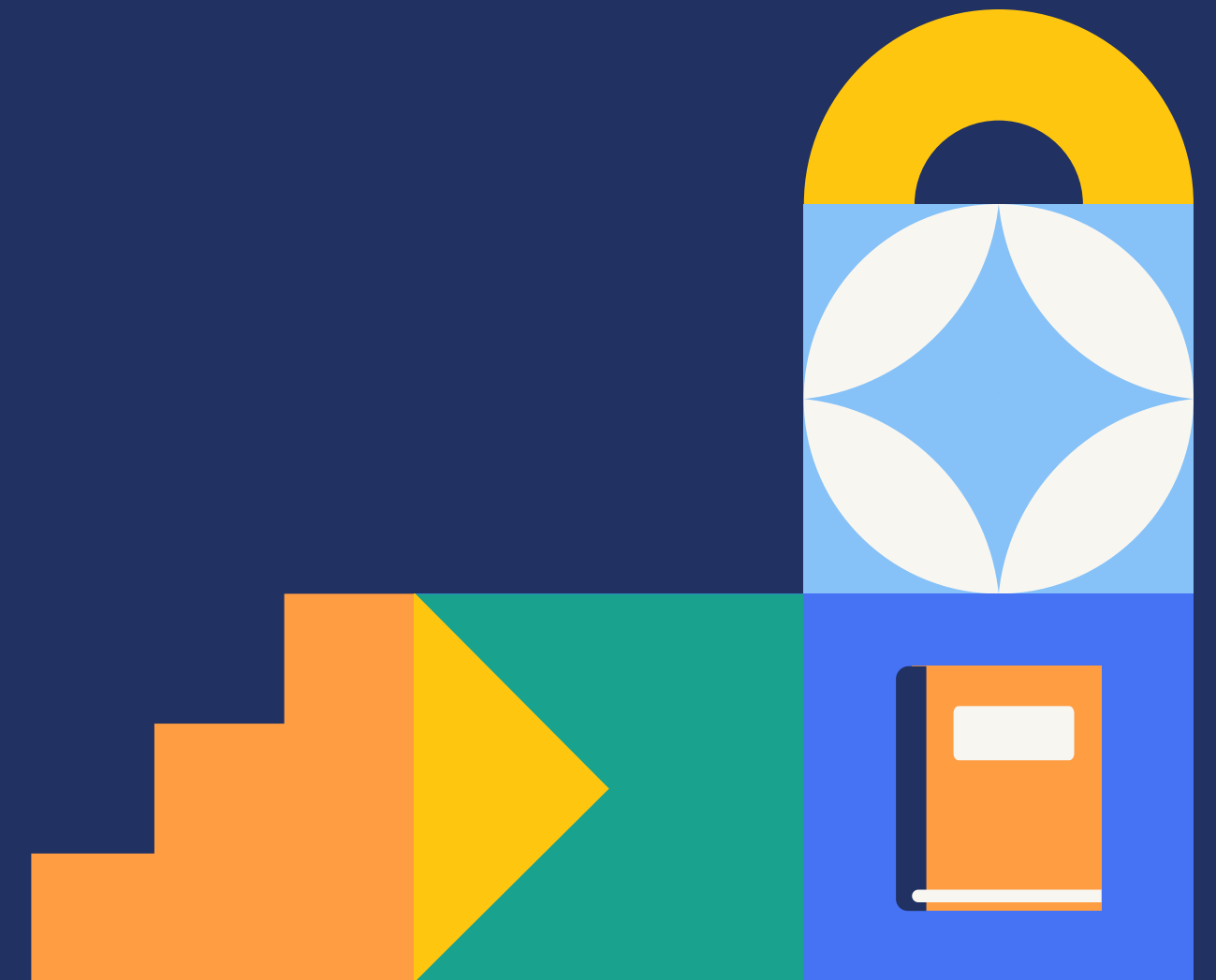
## Exploratory Data Analysis

- **The exploratory data analysis undertaken on this dataset navigates through an extensive process of understanding and visualizing the inherent characteristics of each variable. Descriptive statistics were computed, offering critical insights into central tendencies, variability, and the shape of the distributions. Measures like counts, means, standard deviations, minima, maxima, quartiles, and variability parameters provided a foundation for understanding the dataset's landscape.**
- **Furthermore, visualization techniques such as histograms, scatter plots, and line plots were employed to provide a visual representation of the dataset's trends, distributions, and potential relationships between variables. These visualizations offered a more intuitive understanding of the dataset, helping identify potential correlations, anomalies, or patterns.**



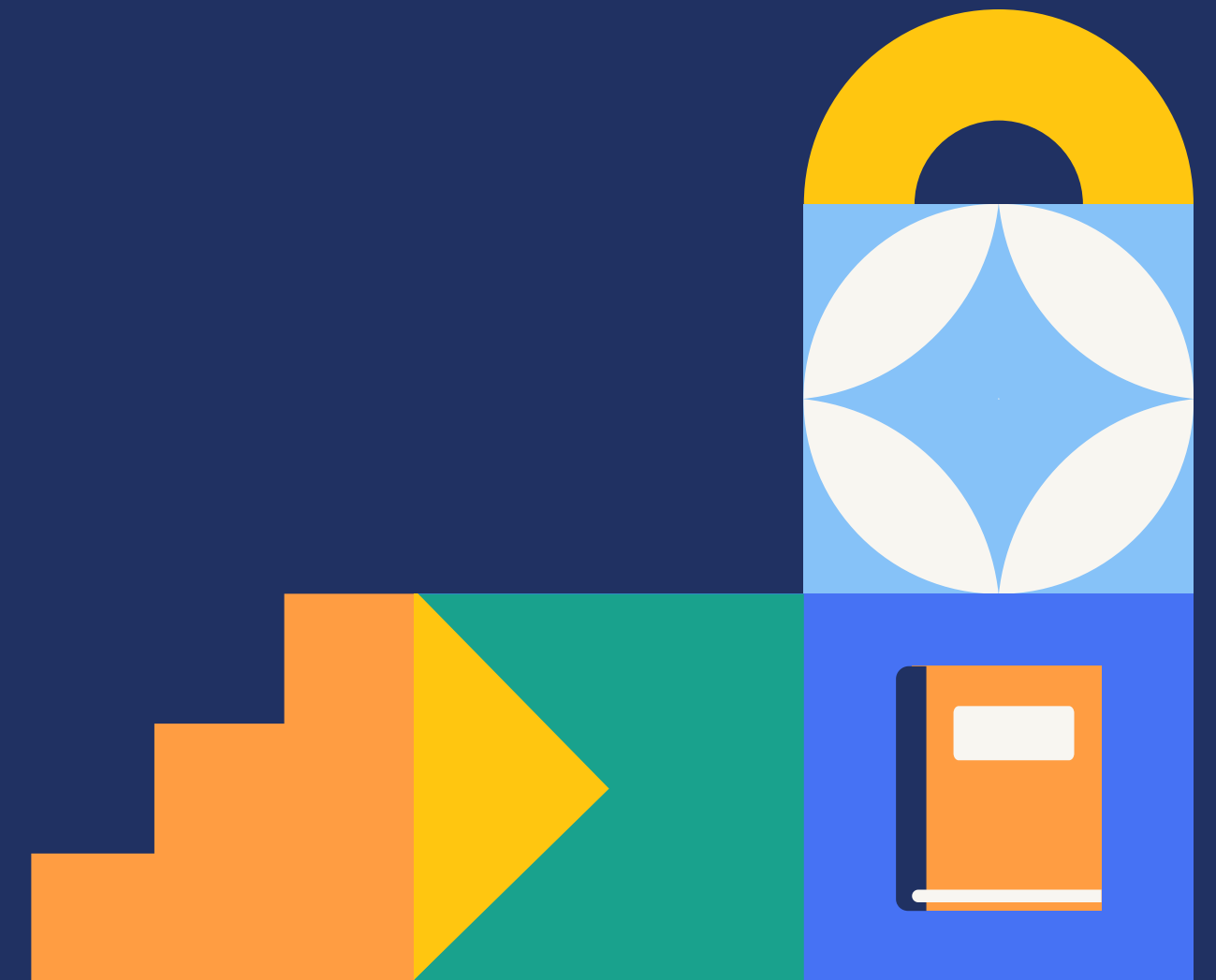
## Scenario 1:- Estimating Total Profit in a Sales Region

- Objective: Estimating the total profit for a new sales region based on historical data.
- Sampling Strategy: Select a random sample of 100 orders from the historical dataset of 5000 orders.
- Statistical Inference: Compute the average total profit and its confidence interval for the sample. Based on this sample, make an inference about the probable range of total profit for the new sales region. For example, if the mean total profit in the sample is 400,000 with a 95 percentage confidence interval of  $\pm 50,000$ , you can infer that the total profit in the new region might be around 400,000 with a 95 percentage probability.



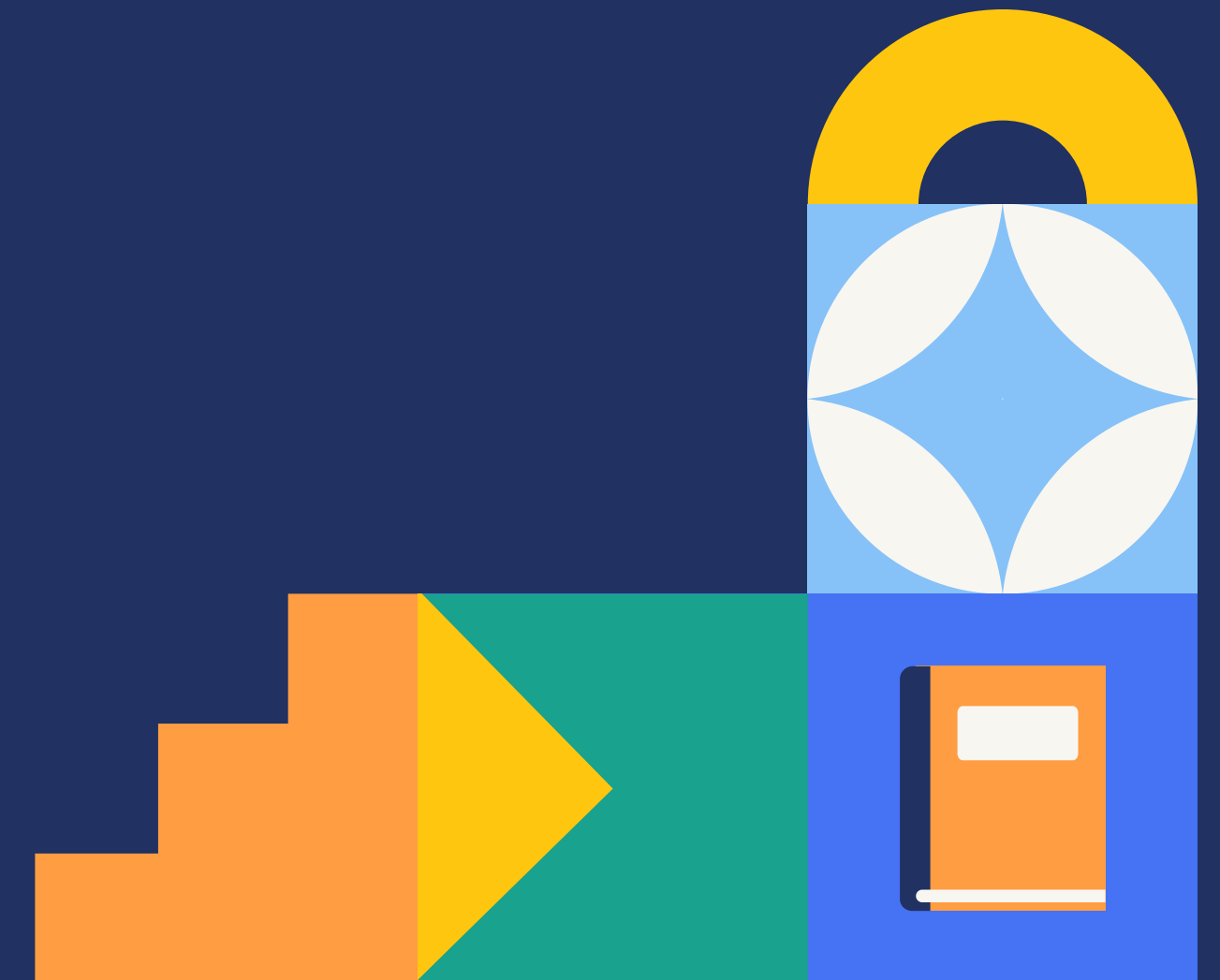
## Scenario 2: Evaluating Unit Price Variability for Item Types

- Objective: Assess the variability in unit prices for different item types.
- Sampling Strategy: Randomly select 3 item types out of the available types and collect unit price data for each type.
- Statistical Inference: Calculate the standard deviation of unit prices for each selected item type. By comparing the standard deviations among these item types, infer which item type has the most or least variability in unit prices. For instance, if the standard deviation for item type A is 50 and for item type B is 100, you can infer that item type B has higher variability in unit prices compared to item type A.



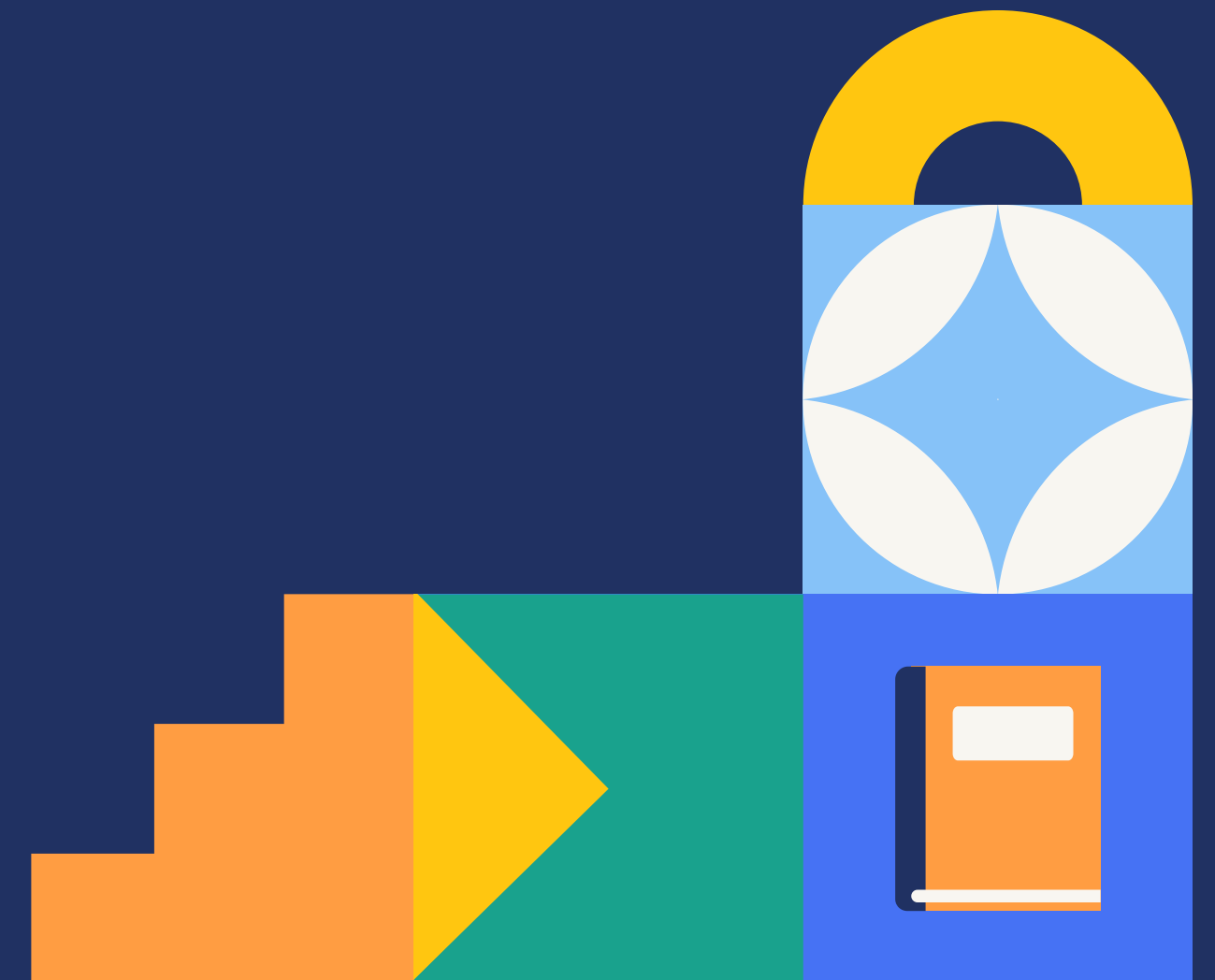
## Scenario 3: Estimating Units Sold for Future Orders

- Objective: Predict the number of units that might be sold for future orders.
- Sampling Strategy: Divide the dataset based on sales channels and select a sample of 50 orders from each channel.
- Statistical Inference: Calculate the mean and standard deviation of units sold for each channel. Based on these sample statistics, infer the probable range of units that might be sold for future orders in each channel. For instance, if the mean units sold in Channel 1 is 4000 with a standard deviation of 500 and in Channel 2 it is 6000 with a standard deviation of 800, you can infer that future orders in Channel 2 might have higher variability and possibly higher units sold compared to Channel 1.



## Scenario 4: Analysing Sales Channel Performance

- Objective: Assess the performance of different sales channels in generating total revenue.
- Sampling Strategy: Randomly select a sample of 200 orders from each sales channel.
- Statistical Inference: Calculate the mean total revenue for each channel in the sample. Based on these means, infer which sales channel is likely to have a higher total revenue. For instance, if the mean revenue in Channel A is \$500,000 and in Channel B is \$600,000, you can infer that Channel B might perform better in generating total revenue.







# Units sold:-

**Count:** There are 5000 data points for Units Sold.

**Mean:** The average number of units sold is approximately 5030.70.

**Standard Deviation:** The data points deviate from the mean by approximately 2914.52 on average.

**Minimum:** The minimum number of units sold is 2.

**25th Percentile (Q1):** 25% of the data lies below 2453.

**Median (50th Percentile):** The middle value is 5123.

**75th Percentile (Q3):** 75% of the data lies below 7576.25.

**Maximum:** The maximum number of units sold is 9999.

**Thank You!**