**1. Frequency Tables**

* **Why?**
  + Helps summarize categorical data by showing the count of occurrences of each category.
  + Used for understanding the distribution of a single categorical variable.
* **Example:**  
  Suppose you have hotel booking data, and you want to see the count of bookings for different meal plans.
* import pandas as pd
* # Sample data
* data = {'type\_of\_meal\_plan': ['Meal Plan 1', 'Meal Plan 2', 'Meal Plan 1', 'Not Selected', 'Meal Plan 2']}
* df = pd.DataFrame(data)
* # Frequency table
* print(df['type\_of\_meal\_plan'].value\_counts())

**Output:**

Meal Plan 1 2

Meal Plan 2 2

Not Selected 1

**2. Two-way Tables**

* **Why?**
  + Shows the relationship between two categorical variables.
  + Helps in understanding how two variables interact.
* **Example:**  
  Suppose you want to see how meal plans are distributed among repeated and non-repeated guests.
* # Sample data
* data = {'type\_of\_meal\_plan': ['Meal Plan 1', 'Meal Plan 2', 'Meal Plan 1', 'Not Selected', 'Meal Plan 2'],
* 'repeated\_guest': [1, 0, 1, 0, 0]}
* df = pd.DataFrame(data)
* # Two-way table
* print(pd.crosstab(df['type\_of\_meal\_plan'], df['repeated\_guest']))

**Output:**

repeated\_guest 0 1

type\_of\_meal\_plan

Meal Plan 1 0 2

Meal Plan 2 2 0

Not Selected 1 0

**3. Two-way Table - Joint Probability**

* **Why?**
  + Joint probability measures the likelihood of two events happening together.
  + Useful for understanding the probability distribution of two categorical variables.
* **Example:**  
  Extending the previous example, we can compute joint probabilities.
* # Joint probability table
* joint\_prob = pd.crosstab(df['type\_of\_meal\_plan'], df['repeated\_guest'], normalize=True)
* print(joint\_prob)

**Output:**

repeated\_guest 0 1

type\_of\_meal\_plan

Meal Plan 1 0.00 0.40

Meal Plan 2 0.40 0.00

Not Selected 0.20 0.00

**4. Two-way Table - Marginal Probability**

* **Why?**
  + Marginal probability is the probability of one event occurring, regardless of the other variable.
  + Helps in analyzing overall probabilities.
* **Example:**  
  Computing marginal probabilities:
* # Marginal probability
* marginal\_prob = joint\_prob.sum(axis=1)
* print(marginal\_prob)

**Output:**

type\_of\_meal\_plan

Meal Plan 1 0.40

Meal Plan 2 0.40

Not Selected 0.20

dtype: float64

**5. Two-way Table - Conditional Probability**

* **Why?**
  + Conditional probability tells us the probability of one event happening given that another event has already happened.
  + Helps in making data-driven decisions.
* **Example:**  
  Probability of repeated guests given a meal plan:
* # Conditional probability (P(A | B))
* conditional\_prob = pd.crosstab(df['type\_of\_meal\_plan'], df['repeated\_guest'], normalize='columns')
* print(conditional\_prob)

**Output:**

repeated\_guest 0 1

type\_of\_meal\_plan

Meal Plan 1 0.00 1.00

Meal Plan 2 1.00 0.00

Not Selected 1.00 0.00

**6. Correlation**

* **Why?**
  + Measures the relationship between two numerical variables.
  + Helps in identifying if an increase in one variable is associated with an increase/decrease in another.
* **Example:**  
  Checking correlation between lead time and price per room:
* # Sample data
* data = {'lead\_time': [20, 50, 100, 200, 10],
* 'avg\_price\_per\_room': [80, 100, 150, 200, 75]}
* df = pd.DataFrame(data)
* # Correlation matrix
* print(df.corr())

**Output:**

lead\_time avg\_price\_per\_room

lead\_time 1.000 0.988

avg\_price\_per\_room 0.988 1.000

* Interpretation:
  + A correlation value close to **1** indicates a strong positive relationship.
  + Here, as **lead time increases, price per room also increases**.