# SAVEETHA SCHOOL OF ENGINEERING

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES SAVEETHA UNIVERSITY**

# Department of Computer Science and Engineering

|  |  |  |
| --- | --- | --- |
| **Course Code: DSA0416**  **(C-Slot)** | **Course Name: Fundamentals of Data Science for Risk Management** | |
| **Branch: AI&DS** | **Year: II/III/IV** | **Date: 26.07.2024** |

1. Imagine you are an analyst for a popular online shopping website. Your task is to analyze customer reviews and provide insights on the average rating and customer satisfaction level for a specific product category. You will use the pandas library to calculate confidence intervals to estimate the true population mean rating. You have been provided with following dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| product\_title | product\_category | star\_rating | review\_headline | review\_date |
| Pineapple  slicer | Apparel | 4 | Really good | "2013-01-14  " |
| Levis Jeans  Pant | Apparel | 5 | Perfect Dress | "2014-04-22  " |
| Wallet | Apparel | 5 | Love it | "2015-07-28  " |
| Salwar | Apparel | 5 | Awesome | "2015-06-12  " |

FREQUENCY DISTRIBUTION

1. A hospital wants to know the most common diseases among their patients. They have a list of all the diseases that their patients have been diagnosed within the past year, along with the number of patients who have been diagnosed with each disease. Write a program that will calculate the frequency distribution of diseases and print out the most common disease using the following dataset

|  |  |
| --- | --- |
| DISEASE\_NAME | IAGNOSED\_PATIENTS |
| Common Cold | 320 |
| Diabetes | 120 |
| Bronchitis | 100 |
| Influenza | 150 |
| Kidney Stones | 60 |

1. A weather station wants to know the most common types of weather in their area. They have a list of all the weather conditions that have occurred in the past year, along with the number of times each weather condition has occurred. Write a program that will calculate the frequency distribution of weather conditions and print out the most common weather type.
2. Scenario: You are a data analyst working for a company that sells products online. You have been tasked with analyzing the sales data for the past month. The data is stored in a Pandas data frame.

Question: Develop a code in Python to find the frequency distribution of the ages of the customers who have made a purchase in the past month.

1. Scenario: You are a data analyst working for a social media platform. As part of your analysis, you have a dataset containing user interaction data, including the number of likes received by each post. Your task is to develop a Python program that calculates the frequency distribution of likes among the posts.

Question: Develop a Python program to calculate the frequency distribution of likes among the posts.

1. Scenario: You are working on a project that involves analyzing customer reviews for a product. You have a dataset containing customer reviews, and your task is to develop a Python program that calculates the frequency distribution of words in the reviews.

Question: Develop a Python program to calculate the frequency distribution of words in the customer review dataset.