

## Lesson 01 Demo 05

### Pattern Identification and Anomaly Detection Using GenAI Tools

**Objective:** To identify patterns and detect anomalies using GenAI tools

**Tools required:** Julius AI

**Prerequisites:** None

#### Steps to be followed:

1. Download the Dataset
2. Upload the Dataset to Julius AI
3. Identify Patterns
4. Detect Anomalies

#### Step 1: Download the Dataset

1.1 Download customer.csv from the reference materials

#### Step 2: Upload the Dataset to Julius AI

#### Step 3: Identify Patterns

3.1 Create a prompt to identify a pattern on the sample dataset uploaded on Julius AI

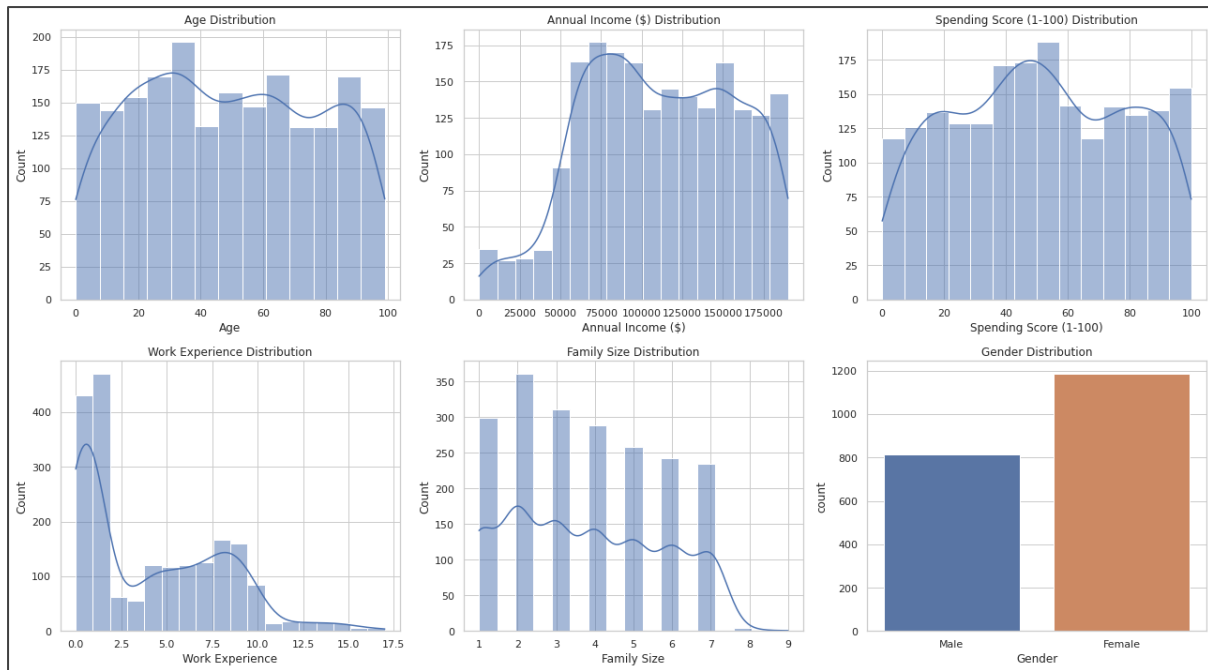
#### Prompt:

Perform pattern identification on this dataset

Let's start with the distribution of numerical features and the gender distribution.

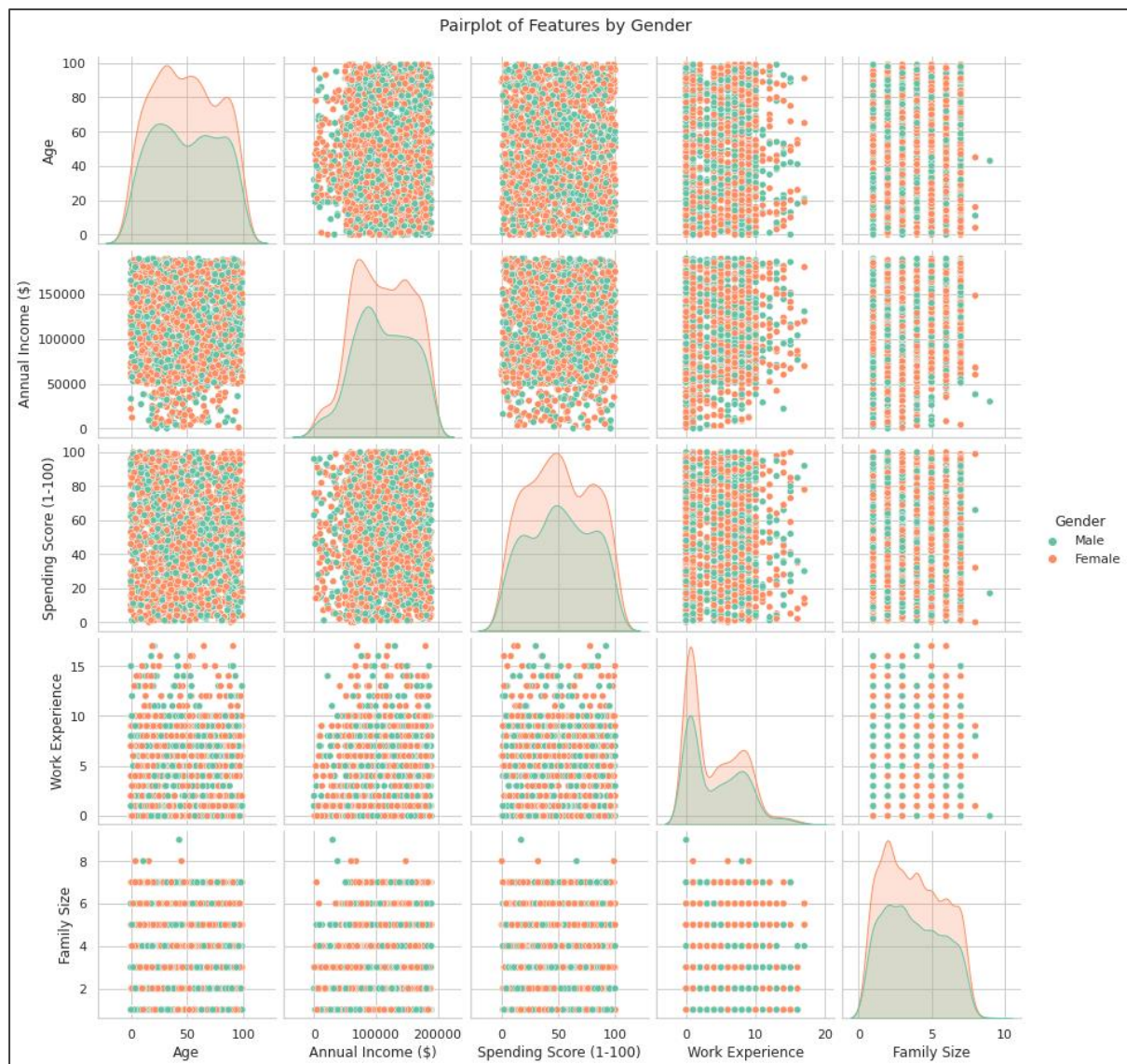
[</> Show Code](#) 

Here are the distributions of the numerical features and the gender distribution in the dataset:



From the distributions, we can observe the following patterns:

- **Age Distribution:** The age of customers is fairly evenly distributed, with a slight increase in the younger age groups.
- **Annual Income Distribution:** Most customers have an annual income in the lower to middle range, with fewer customers having very high incomes.
- **Spending Score Distribution:** The spending score is fairly evenly distributed across all scores, indicating a variety of spending behaviors among customers.
- **Work Experience Distribution:** A significant number of customers have low to moderate work experience, with fewer customers having high levels of experience.
- **Family Size Distribution:** Most customers have small to medium-sized families, with fewer customers having very large families.
- **Gender Distribution:** The dataset contains a relatively balanced distribution of male and female customers.



This pairplot allows us to observe several relationships and distributions within the dataset, segmented by gender. Here are some key observations:

- **Age and Spending Score:** There doesn't appear to be a strong linear relationship between age and spending score across genders. However, younger customers tend to have a wider range of spending scores.
- **Annual Income and Spending Score:** There's a notable cluster of customers with moderate annual incomes and high spending scores, suggesting a segment of the market that, despite not having the highest incomes, is willing to spend significantly. This pattern is visible across both genders.
- **Work Experience:** The distribution of work experience relative to other features such as age, annual income, and spending score does not show a clear pattern, indicating that work experience might not directly influence spending behavior or income levels in this dataset.
- **Family Size:** Similar to work experience, family size does not show a clear relationship with spending score or annual income. However, there's a slight indication that larger family sizes might correlate with lower spending scores, which could suggest budget constraints.

These insights can help in identifying customer segments and tailoring marketing strategies accordingly.

Step 4: Detect Anomalies

Prompt:

Identify customers who are deviating from their usual spending patterns

J

</> Show Code

(Optional) click on columns, rows, or cells to focus on specific data

	CustomerID	Gender	Age	Annual Income (\$)	Spending Score (1-100)	Profession	Work Experience	Family Size	Anomaly
0	11	Male	67	7000	14	Engineer	1	3	Anomaly
1	23	Female	46	42000	5	Artist	13	2	Anomaly
2	34	Male	18	62000	92	Homemaker	9	7	Anomaly
3	43	Male	48	22000	36	Artist	14	3	Anomaly
4	48	Female	27	71000	47	Healthcare	12	1	Anomaly

The model identified several customers as anomalies based on their spending patterns and other features such as age, annual income, work experience, and family size. These customers deviate significantly from the typical patterns observed in the dataset. For instance, one anomaly is a 67-year-old male with an annual income of \$7,000 and a spending score of 14, which might indicate unusually low spending relative to his age group. Another example is an 18-year-old male with a high annual income of \$62,000 and a very high spending score of 92, suggesting unusually high spending for his age and income level. These anomalies could represent unique customer behaviors or data entry errors and may warrant further investigation.