AI In Personalized Advertisement - Ethics & Manipulation

March 3, 2025

1 By Hussain B.

www.linkedin.com/in/hussainbagasrawala

www.github.com/ihussainb/AI-Ethics-Project

```python
[92]: import pandas as pd
      import matplotlib.pyplot as plt
      import seaborn as sns
      import numpy as np
      from wordcloud import WordCloud
      from scipy.stats import chi2_contingency
      from sklearn.decomposition import PCA
      from sklearn.preprocessing import StandardScaler
      from sklearn.cluster import KMeans, AgglomerativeClustering
      from sklearn.metrics import silhouette_score
      from scipy.cluster.hierarchy import dendrogram, linkage
      from mlxtend.frequent_patterns import apriori, association_rules, fpgrowth
      from mlxtend.preprocessing import TransactionEncoder

      import warnings
      warnings.filterwarnings("ignore", category=FutureWarning)
```

```python
[2]: # Load the dataset
     file_path = "Survey on AI and Personalized Advertising Ethics (Responses) -↵
      ↪Form responses 1.csv"
     df = pd.read_csv(file_path)
```

```python
[3]: # Rename columns to correct extra spaces
     df.columns = [col.strip() for col in df.columns]
```

```python
[4]: # Set style
     sns.set_style("darkgrid")
     plt.rcParams.update({'font.size': 12, 'axes.titlesize': 16, 'axes.labelsize':↵
      ↪14})
```

```python
[5]: # 1. Distribution of Age Groups
     plt.figure(figsize=(9, 5))
```

```
sns.countplot(x=df["Q1: What is your age group?"],
              order=df["Q1: What is your age group?"].value_counts().index,
              palette="viridis")
plt.title("Age Group Distribution")
plt.xlabel("Age Group")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.show()
```


Age Group Distribution
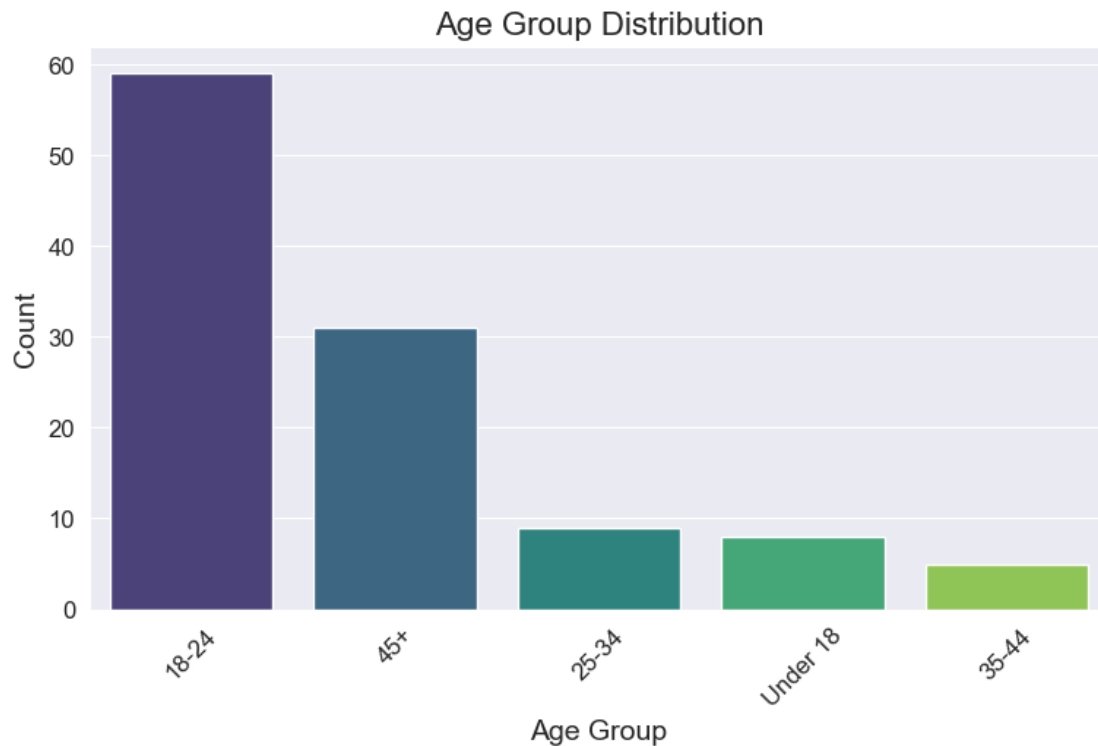
This chart shows the distribution of respondents by age group. It helps identify which age group is most represented in the survey and whether the responses are biased toward a particular demographic.

```
[6]: # 2. Occupation Distribution
plt.figure(figsize=(9, 5))
sns.countplot(y=df["Q2: What is your occupation?"],
              order=df["Q2: What is your occupation?"].value_counts().index,
              palette="magma")
plt.title("Occupation Distribution")
plt.xlabel("Count")
plt.ylabel("Occupation")
plt.show()
```

Occupation Distribution
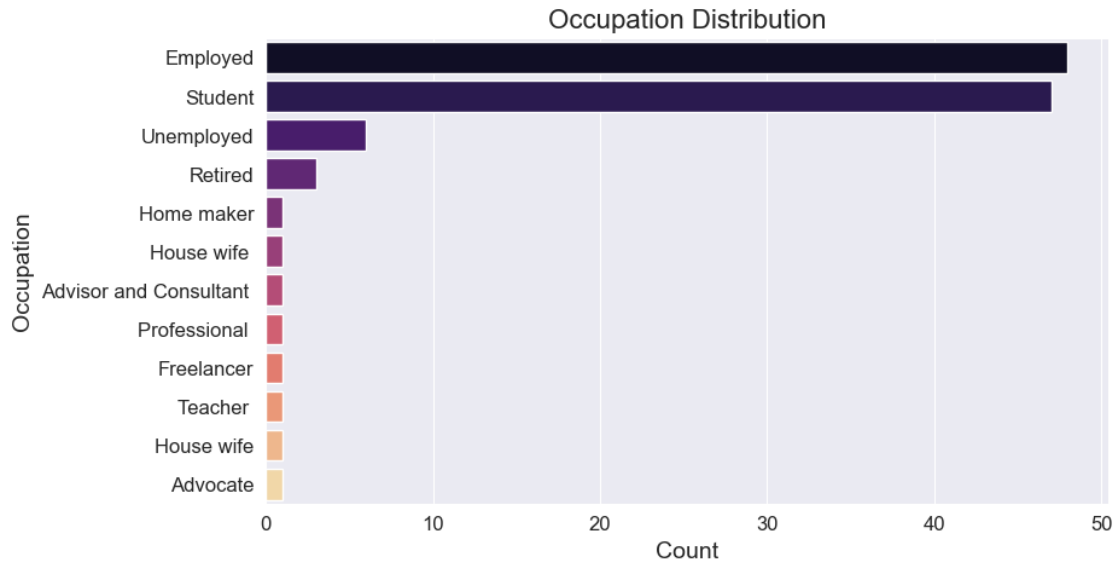
The occupation distribution visualizes the professional background of respondents. This helps in understanding if specific occupations influence perspectives on AI and personalized ads.

```
[8]:  # 3. Concern Level About Personal Data Collection (Box Plot)
      plt.figure(figsize=(7, 5))
      sns.boxplot(y=df["Q7: How concerned are you about the amount of personal data␣
       ↪companies collect to show personalized ads?"],
                  color="dodgerblue")
      plt.title("Concern About Personal Data Collection")
      plt.ylabel("Concern Level (1-5)")
      plt.show()
```
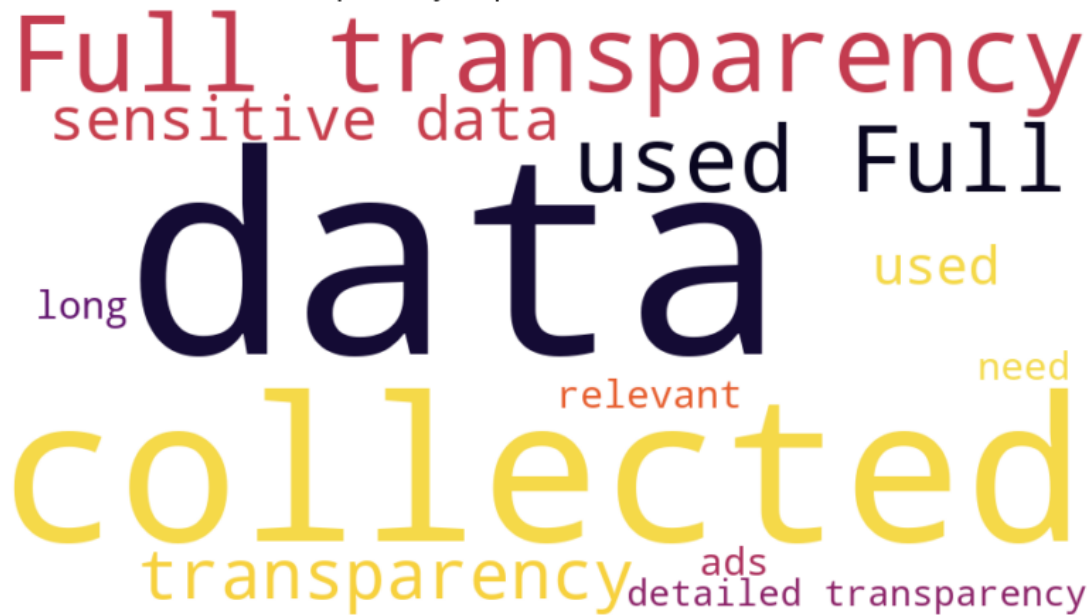
Concern About Personal Data Collection
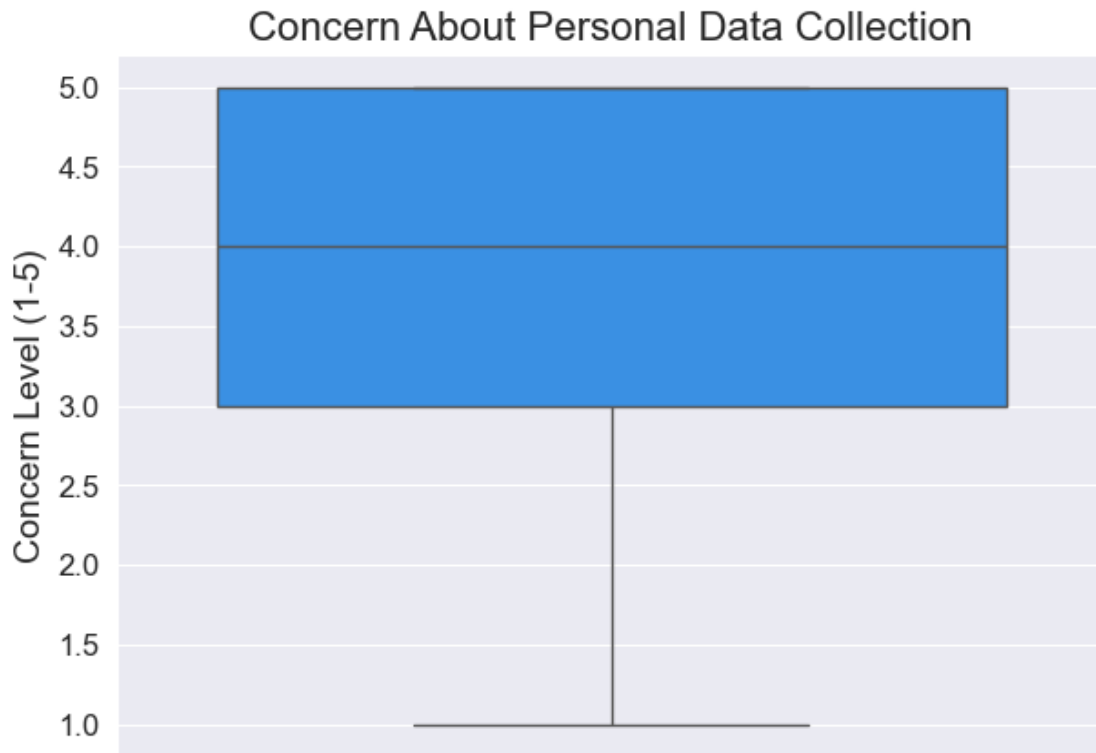

```
[9]: # 4. Word Cloud of Transparency Expectations
     text_data = " ".join(df["Q10: What level of transparency do you expect from␣
      ↪companies that use AI for personalized advertising?"].dropna().astype(str))
     wordcloud = WordCloud(width=900, height=500, background_color="white",␣
      ↪colormap="inferno").generate(text_data)

     plt.figure(figsize=(10, 6))
     plt.imshow(wordcloud, interpolation="bilinear")
     plt.axis("off")
     plt.title("Transparency Expectations - Word Cloud")
     plt.show()
```

Transparency Expectations - Word Cloud

Full transparency
sensitive data
used Full
data
long
used
relevant
need
collected
transparency
ads
detailed transparency

```
[10]: # 5. Sentiment Distribution on Personalized Ads
      plt.figure(figsize=(9, 5))
      sns.countplot(x=df["Q9: When encountering personalized ads, do you feel you've␣
       ↪given sufficient consent for your data to be used?"],
                   order=df["Q9: When encountering personalized ads, do you feel␣
       ↪you've given sufficient consent for your data to be used?"].value_counts().
       ↪index,
                   palette="coolwarm")
      plt.title("Sentiment Distribution on Personalized Ads")
      plt.xlabel("Sentiment")
      plt.ylabel("Count")
      plt.xticks(rotation=45)
      plt.show()
```

Sentiment Distribution on Personalized Ads
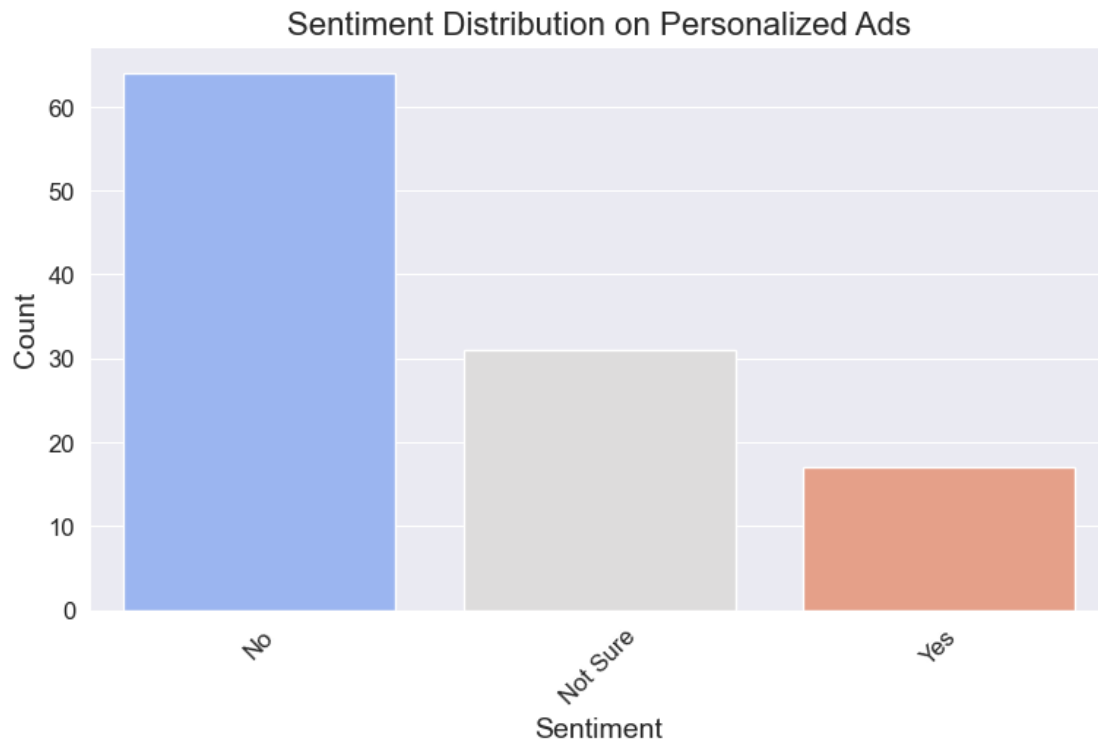
```
[11]:  # 6. Pie Chart for Trust in AI-Powered Ads
       plt.figure(figsize=(8, 8))
       df["Q12: Do you think personalized ads are more persuasive than traditional ads?
        ↪"].value_counts().plot.pie(
           autopct="%1.1f%%", colors=["#ff9999", "#66b3ff", "#99ff99"],↵
        ↪startangle=140, explode=[0.05, 0.05, 0.05])
       plt.title("Trust in AI-Powered Personalized Ads")
       plt.ylabel("")
       plt.show()
```

Trust in AI-Powered Personalized Ads

No, they are less influential

No, they are about the same

6.2%

33.0%

60.7%

Yes, they influence my decisions more
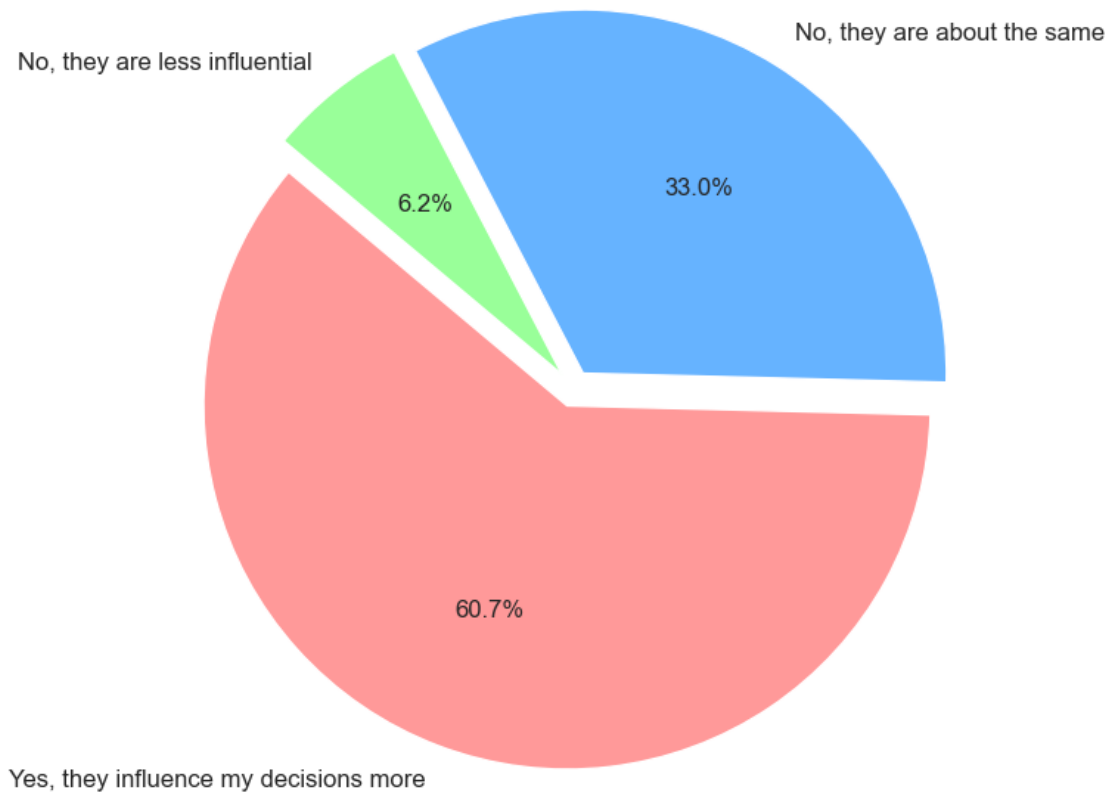
[12]:
```python
# 7. Relationship Between Age and Trust in AI-Powered Ads
plt.figure(figsize=(10, 6))
sns.countplot(x=df["Q1: What is your age group?"], hue=df["Q12: Do you think␣
 ↪personalized ads are more persuasive than traditional ads?"],␣
 ↪palette="Spectral")
plt.title("Trust in AI-Powered Ads by Age Group")
plt.xlabel("Age Group")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.legend(title="Trust Level")
plt.show()
```

Trust in AI-Powered Ads by Age Group
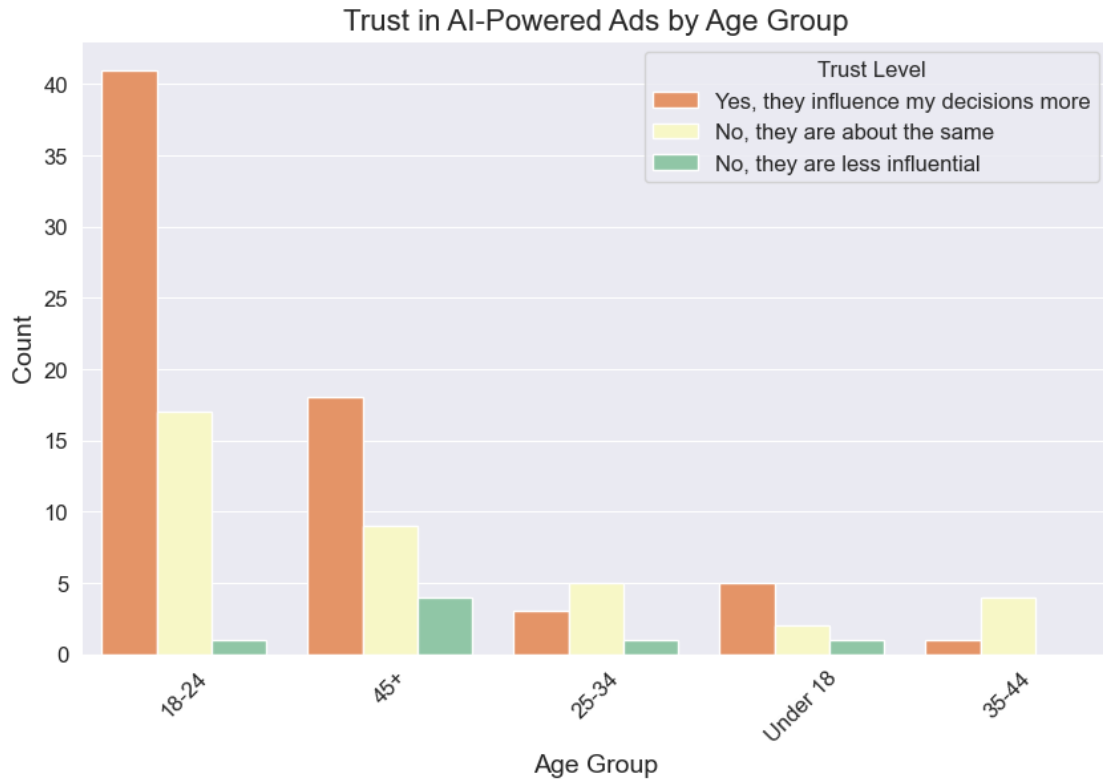

```
[13]: # 8. Bar Chart for Most Used Ad Platforms
      plt.figure(figsize=(10, 5))
      platform_counts = df["Q5: Which platforms do you see personalized ads on the␣
        ↪most? (Select all that apply)"].str.split(', ').explode().value_counts()
      platform_counts.plot(kind="bar", color="royalblue", edgecolor="black")
      plt.title("Most Used Platforms for Personalized Ads")
      plt.xlabel("Platform")
      plt.ylabel("Count")
      plt.xticks(rotation=45)
      plt.show()
```

Most Used Platforms for Personalized Ads
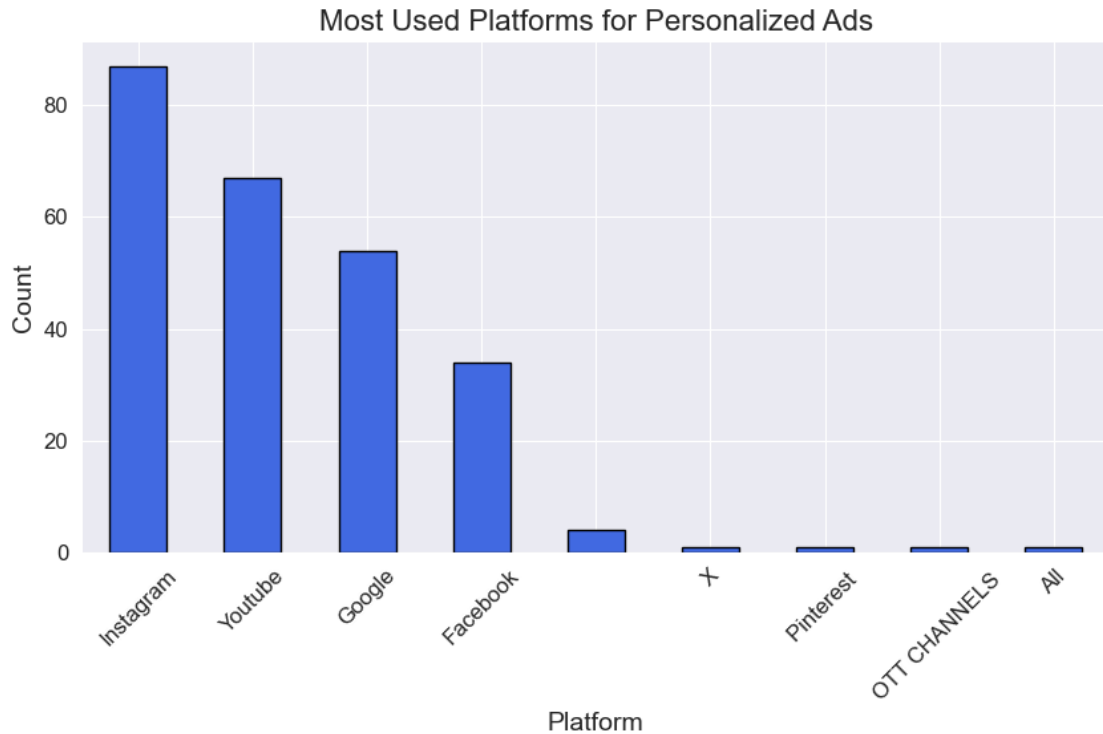
2 Hypothesis Testings

```
[18]: #Getting the data ready for hypothesis testings by converting appropriate␣
      ↪categorical columns to numeric
```

```
[19]: # Strip extra spaces from column names
      df.columns = df.columns.str.strip()
```

```
[39]: # Selecting relevant columns for hypothesis testing
      columns_to_keep = [
          "Q7: How concerned are you about the amount of personal data companies␣
      ↪collect to show personalized ads?",  # Privacy
          "Q10: What level of transparency do you expect from companies that use AI␣
      ↪for personalized advertising?",  # Transparency
          "Q14: Have you noticed that personalized ads are biased toward targeting␣
      ↪specific demographic groups?",  # Fairness
          "Q15: Do you think incorporating bias detection mechanisms in AI systems␣
      ↪can improve fairness in personalized ads?",  # Bias Mitigation
          'Q11: To what extent do you agree with the following statement:␣
      ↪"Personalized ads can manipulate my purchasing decisions by exploiting my␣
      ↪preferences or emotions."',  # Manipulation
```

```
      "Q9: When encountering personalized ads, do you feel you've given␣
    ↪sufficient consent for your data to be used?",   # Consent
      "Q19: Do you feel that you have adequate control over the types of␣
    ↪personalized ads you see?"   # User Control / Trust
]
```

[40]:
```
# Filter dataset with required columns
df_hypothesis = df[columns_to_keep].dropna()
```

[41]:
```
# Convert categorical responses into numerical codes for statistical testing
def encode_categories(series):
    return series.astype("category").cat.codes
```

[42]:
```
df_hypothesis = df_hypothesis.apply(encode_categories)
```

[43]:
```
# Chi-Square test
def chi_square_test(*cols):
    contingency_table = pd.crosstab(index=df_hypothesis[cols[0]],␣
  ↪columns=[df_hypothesis[col] for col in cols[1:]])
    chi2, p, dof, expected = chi2_contingency(contingency_table)
    return chi2, p
```

2.0.1 Hypothesis 1 - "The use of AI in personalized advertising leads to significant ethical concerns regarding privacy, fairness, and transparency compared to traditional advertising methods."

[44]:
```
chi2_1, p_1 = chi_square_test("Q7: How concerned are you about the amount of␣
    ↪personal data companies collect to show personalized ads?",
                             "Q10: What level of transparency do you expect␣
    ↪from companies that use AI for personalized advertising?",
                             "Q14: Have you noticed that personalized ads are␣
    ↪biased toward targeting specific demographic groups?")
```

[45]:
```
print(f"Hypothesis 1: Privacy, Fairness, Transparency -> Chi-Square = {chi2_1:.
    ↪4f}, p-value = {p_1:.4f}")
```

```
Hypothesis 1: Privacy, Fairness, Transparency -> Chi-Square = 46.6460, p-value =
0.0149
```

This p-value is below 0.05, indicating statistical significance. This means that there is a significant relationship between privacy concerns, fairness (bias concerns), and transparency expectations in AI-driven personalized advertising. Interpretation: Users who express strong concerns about data privacy also tend to have concerns about bias in ads and expect higher transparency from companies using AI for advertising. Conclusion: The hypothesis is supported—AI-based personalized ads raise ethical concerns in multiple dimensions compared to traditional advertising.

2.0.2 Hypothesis 2 - "The use of AI in personalized advertising leads to significant ethical concerns regarding privacy, fairness, and transparency compared to traditional advertising methods."

```
[46]: chi2_2, p_2 = chi_square_test("Q15: Do you think incorporating bias detection␣
      ↪mechanisms in AI systems can improve fairness in personalized ads?",
                                      'Q11: To what extent do you agree with the␣
      ↪following statement: "Personalized ads can manipulate my purchasing␣
      ↪decisions by exploiting my preferences or emotions."')
```

```
[47]: print(f"Hypothesis 2: Bias Detection vs. Manipulation -> Chi-Square = {chi2_2:.
      ↪4f}, p-value = {p_2:.4f}")
```

```
Hypothesis 2: Bias Detection vs. Manipulation -> Chi-Square = 19.0318, p-value =
0.0147
```

Since the p-value is below 0.05, the relationship between perceptions of bias detection tools and manipulation concerns is statistically significant. Interpretation: People who believe that bias detection tools can improve fairness in AI-driven advertising are also more likely to have concerns about manipulative advertising practices. Conclusion: The hypothesis is supported—users perceive bias mitigation as an important factor in reducing manipulation in AI-driven advertising.

2.0.3 Hypothesis 3: "Increasing transparency and simplifying consent mechanisms will lead to higher user trust and acceptance of personalized advertising."

```
[56]: chi2_3, p_3 = chi_square_test("Q10: What level of transparency do you expect␣
      ↪from companies that use AI for personalized advertising?",
                                     "Q9: When encountering personalized ads, do you␣
      ↪feel you've given sufficient consent for your data to be used?")
      print(f"Hypothesis 3a: Transparency vs. Consent -> Chi-Square = {chi2_3:.4f},␣
      ↪p-value = {p_3:.4f}")
```

```
Hypothesis 3a: Transparency vs. Consent -> Chi-Square = 14.5981, p-value =
0.0056
```

Since the p-value is below 0.05 it is significant. Interpretation: Higher transparency expectations are linked to a stronger perception that users should be given more control over their data and consent.

```
[57]: chi2_3b, p_3b = chi_square_test("Q10: What level of transparency do you expect␣
      ↪from companies that use AI for personalized advertising?",
                                       "Q19: Do you feel that you have adequate␣
      ↪control over the types of personalized ads you see?")
      print(f"Hypothesis 3b: Transparency vs. User Control -> Chi-Square = {chi2_3b:.
      ↪4f}, p-value = {p_3b:.4f}")
```

```
Hypothesis 3b: Transparency vs. User Control -> Chi-Square = 9.4506, p-value =
0.1498
```

The p-value here is greater than 0.05 which is not significant. While transparency expectations are

important, they do not necessarily translate into people feeling they have more control over the ads they see.

3 Principal Component Analysis

I conducted hypothesis testing and found relationships between privacy, fairness, transparency, bias detection, manipulation, consent, and user control. PCA can be applied to see if these patterns naturally emerge from the data. It helps check if survey responses can be grouped into broader latent themes rather than treating each question separately.

```python
[62]: # Filter dataset with required columns
      df_analysis = df[columns_to_keep].dropna()
```

```python
[63]: # Convert categorical responses into numerical codes for statistical testing
      def encode_categories(series):
          return series.astype("category").cat.codes
```

```python
[64]: df_analysis = df_analysis.apply(encode_categories)
```

```python
[65]: # Standardize the data
      scaler = StandardScaler()
      df_scaled = scaler.fit_transform(df_analysis)
```

```python
[66]: # Perform PCA with optimal number of components
      pca = PCA(n_components=3)   # Extract 3 main components
      principal_components = pca.fit_transform(df_scaled)
```

```python
[67]: pca_df = pd.DataFrame(principal_components, columns=["Component 1", "Component␣
      ↪2", "Component 3"])
```

```python
[68]: explained_variance = pca.explained_variance_ratio_
      print(f"Explained Variance: {explained_variance}")
```

```
Explained Variance: [0.27043551 0.1785564  0.13433874]
```

This tells us how much of the total variance (spread of data) is captured by each principal component.

Component 1 explains 27.04% of the variance. Component 2 explains 17.85% of the variance. Component 3 explains 13.43% of the variance.

Together, these three components explain around 58.3% of the total variance in the dataset.

```python
[69]: oadings = pd.DataFrame(pca.components_.T, index=df_analysis.columns,␣
      ↪columns=["Component 1", "Component 2", "Component 3"])
      print("\nPCA Loadings:")
      print(loadings)
```

```
PCA Loadings:
```

```
                                                    Component 1  Component 2  \
Q7: How concerned are you about the amount of p…      0.535986     0.186157
Q10: What level of transparency do you expect f…    -0.374926     0.247422
Q14: Have you noticed that personalized ads are…     0.026044     0.557238
Q15: Do you think incorporating bias detection …     0.366660     0.292895
Q11: To what extent do you agree with the follo…     0.468451     0.101420
Q9: When encountering personalized ads, do you …   -0.461522     0.287819
Q19: Do you feel that you have adequate control…    -0.067695     0.643972


                                                    Component 3
Q7: How concerned are you about the amount of p…     -0.129878
Q10: What level of transparency do you expect f…      0.409675
Q14: Have you noticed that personalized ads are…      0.646490
Q15: Do you think incorporating bias detection …     -0.080149
Q11: To what extent do you agree with the follo…      0.142062
Q9: When encountering personalized ads, do you …    -0.287946
Q19: Do you feel that you have adequate control…     -0.536498
```

Values close to +1 or -1 mean that the question is strongly related to that principal component.

Values near 0 mean little to no contribution to that component

```python
[70]:  plt.figure(figsize=(8, 6))
       sns.heatmap(loadings, annot=True, cmap="coolwarm")
       plt.title("PCA Loadings: Ethical Concerns in AI-Based Personalized Ads")
       plt.show()
```


PCA Loadings: Ethical Concerns in AI-Based Personalized Ads
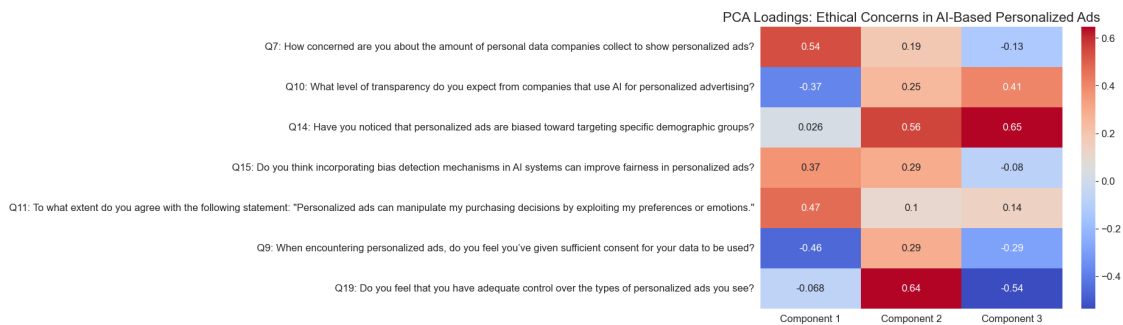
```python
[71]:  eigenvalues = pca.explained_variance_
       plt.figure(figsize=(8, 5))
       plt.bar(range(1, len(explained_variance) + 1), explained_variance * 100)
       plt.xlabel("Principal Component")
       plt.ylabel("Variance Explained (%)")
       plt.title("Explained Variance by PCA Components")
       plt.show()
```

Explained Variance by PCA Components
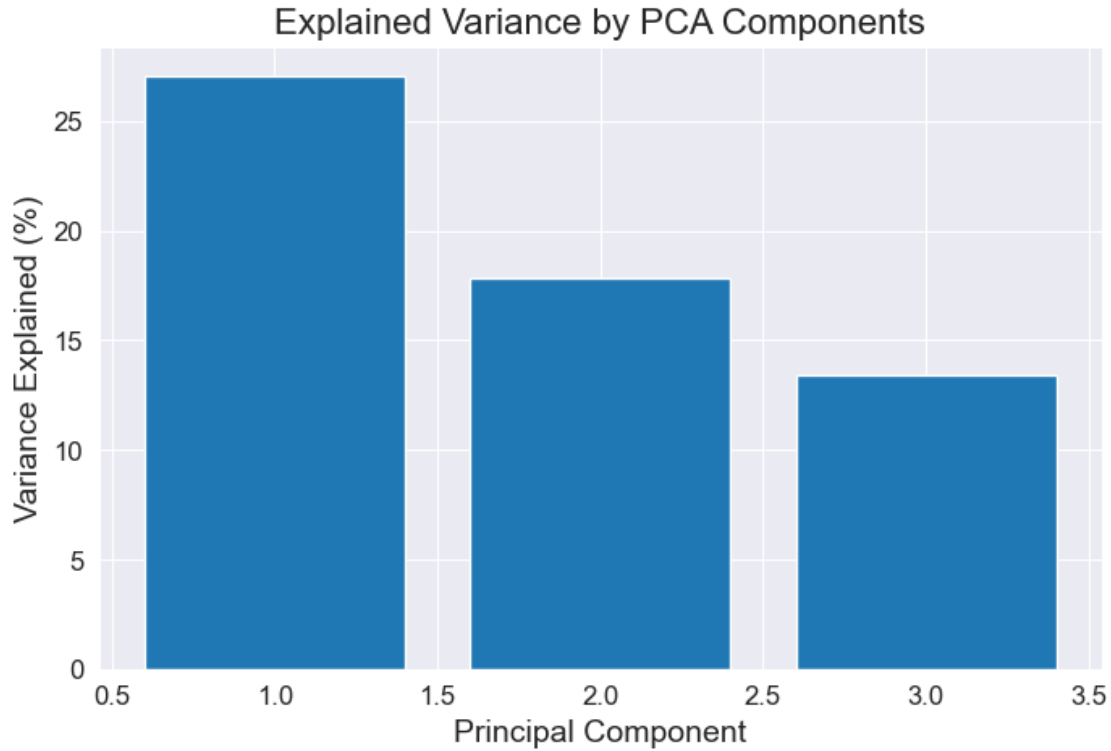
3.0.1 Component 1 (Privacy, Transparency, and Fairness)

High positive loadings: Privacy Concerns (0.5359) Manipulation Concerns (0.4684) Bias Detection (0.3666)

What it means:

This component represents concerns related to data privacy, bias in advertising, and manipulation through AI-driven ads. Users who strongly worry about privacy also tend to believe that bias detection tools matter and that personalized ads can manipulate them. Negative loading on transparency (-0.3749) means that people who expect high transparency tend to worry less about privacy issues.

Conclusion: This suggests that privacy concerns, fairness concerns, and transparency expectations are linked—aligning with first hypothesis.

3.0.2 Component 2 (Bias Detection and User Control)

High positive loadings: Fairness (Bias in Ads) (0.5572) User Control over Ads (0.6439)

What it means:

This component represents attitudes towards bias in AI advertising and whether people feel in control. People who believe AI-driven ads are biased toward specific groups also feel that they lack control over the ads they see. Bias detection mechanisms (0.2928) also contribute here, meaning that belief in bias mitigation is somewhat linked to control.

Conclusion: This supports the idea that bias concerns and lack of control over personalized ads are closely related—which aligns with parts of the second and third hypotheses.

3.0.3 Component 3 (Transparency, Fairness, and Control)

High positive loadings: Transparency Expectations (0.4096) Fairness (Bias in Ads) (0.6464)

What it means:

People who expect high transparency tend to also perceive bias in ads. Negative loading on user control (-0.5364) suggests that those who want more transparency don't necessarily feel in control over their ads. Negative loading on consent (-0.2879) suggests that consent expectations are weakly linked to transparency.

Conclusion: This partially supports your third hypothesis, showing that transparency and fairness concerns are interlinked, but transparency does not necessarily mean people feel in control.

4 K-Means Clustering

Clustering is an unsupervised learning technique used to group similar data points together. Since we don't have predefined groups (labels), we use clustering to discover patterns among respondents.

For this study, clustering helps us identify distinct respondent groups based on their ethical concerns regarding AI-driven personalized advertising.

K-Means is chosen because:

It efficiently partitions respondents into K clusters (K=3 here). It minimizes intra-cluster variance, ensuring respondents within the same cluster are similar. It works well with PCA-reduced data, making clusters more distinct.

Before clustering, we applied Principal Component Analysis (PCA) to reduce dimensionality. Instead of using all original survey responses, we projected the data onto three principal components (PC1, PC2, PC3) that capture the most variance.

This ensures:

1) Better clustering separation

2) Less noise from redundant information

3) Easier visualization of group differences

```python
[78]: # Filter dataset with required columns
      df_cluster = df[columns_to_keep].dropna()
```

```python
[79]: # Convert categorical responses into numerical codes
      def encode_categories(series):
          return series.astype("category").cat.codes
```

```python
[80]: df_cluster = df_cluster.apply(encode_categories)
```

```
[81]: scaler = StandardScaler()
      df_scaled = scaler.fit_transform(df_cluster)
```

```
[82]: pca = PCA(n_components=3)
      principal_components = pca.fit_transform(df_scaled)
```

```
[83]: wcss = []   # Within-cluster sum of squares
      for i in range(1, 11):
          kmeans = KMeans(n_clusters=i, random_state=42, n_init=10)
          kmeans.fit(principal_components)
          wcss.append(kmeans.inertia_)
```

```
[84]: plt.figure(figsize=(8, 5))
      plt.plot(range(1, 11), wcss, marker='o', linestyle='--')
      plt.xlabel('Number of Clusters')
      plt.ylabel('WCSS')
      plt.title('Elbow Method to Determine Optimal K')
      plt.show()
```
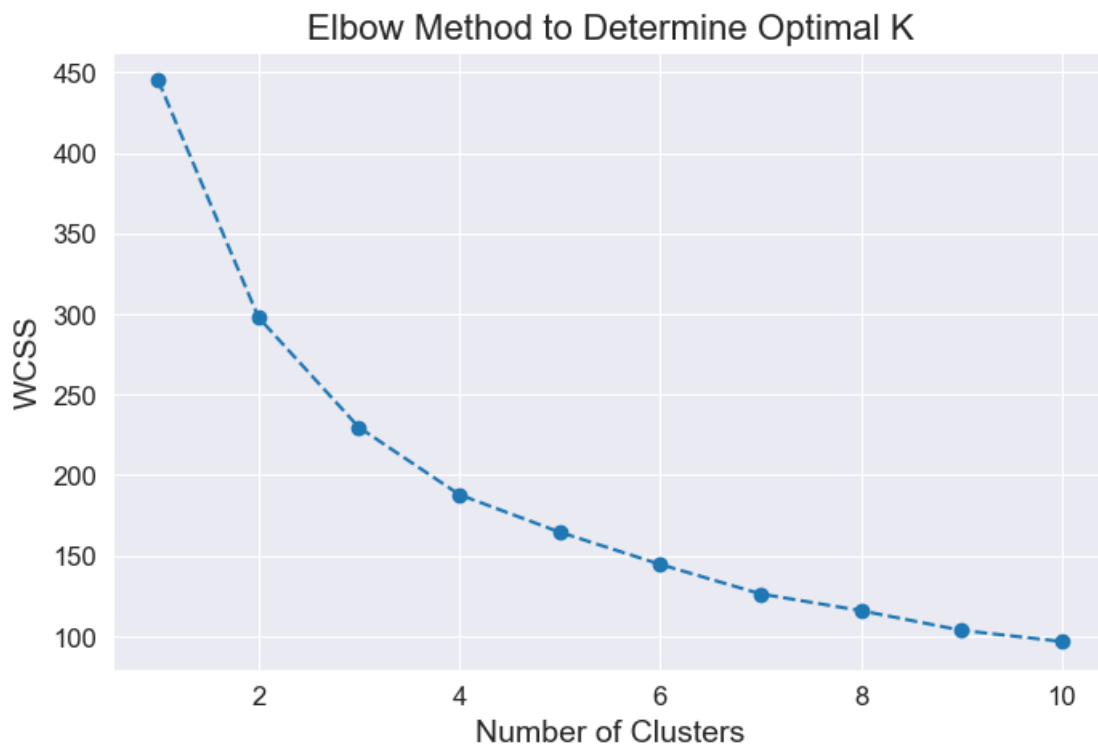


```
[85]: kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
      clusters = kmeans.fit_predict(principal_components)
```

```
[86]: df_cluster['Cluster'] = clusters
```

```
[87]: plt.figure(figsize=(8, 6))
      sns.scatterplot(x=principal_components[:, 0], y=principal_components[:, 1],␣
        ↪hue=clusters, palette='viridis')
      plt.xlabel('Principal Component 1')
      plt.ylabel('Principal Component 2')
      plt.title('K-Means Clustering of Respondents')
      plt.show()
```

K-Means Clustering of Respondents
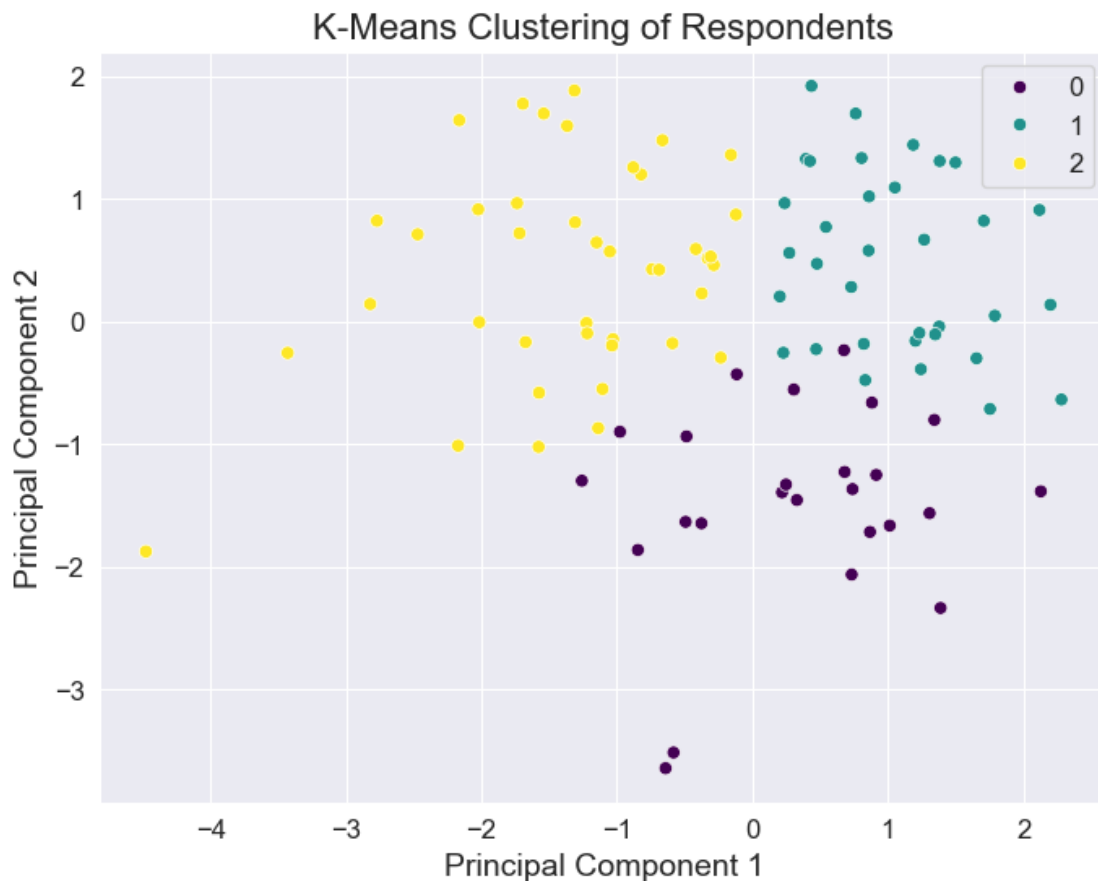

```
[88]: print("Cluster Centers:")
      print(pd.DataFrame(kmeans.cluster_centers_, columns=['PC1', 'PC2', 'PC3']))
```

```
Cluster Centers:
        PC1       PC2       PC3
0  0.318949 -1.406559 -0.445755
1  1.201303  0.527425  0.104520
2 -1.349137  0.401907  0.187015
```

Each cluster center represents the average position of all respondents in that group.

4.0.1 PC1 (Privacy vs. Transparency Concerns):

Positive values (Cluster 1) → More privacy concern, lower transparency satisfaction

Negative values (Cluster 2) → Less privacy concern, higher transparency satisfaction

4.0.2 PC2 (Bias Awareness vs. Trust in AI Ads):

Negative values (Cluster 0) → Higher trust, lower bias concerns

Positive values (Cluster 2) → Higher bias concerns

4.0.3 PC3 (User Control & Manipulation Perception):

Negative values (Cluster 2) → Feel manipulated, lack control

Positive values (Cluster 1) → Feel in control, less concern

4.0.4 Cluster 0: Privacy-Conscious but Low Bias Awareness

1) High privacy concern (Q7)

2) Low expectation for transparency (Q10)

3) Do NOT perceive bias in ads (Q14)

4) Moderate concern about manipulation (Q11)

5) Feel they lack control over ads (Q19)

Interpretation: This group cares a lot about privacy but isn't very aware of bias issues in AI-driven ads. They feel like they lack control over ads, yet they aren't actively pushing for transparency.

4.0.5 Cluster 1: Strong Ethical Concern Advocates

1) Highest privacy concern (Q7)

2) Very high bias awareness (Q14)

3) Strong belief in bias detection as a solution (Q15)

4) Perceive manipulation in personalized ads (Q11)

5) Feel they lack consent over data usage (Q9)

6) Demand higher transparency (Q10)

Interpretation: This cluster actively opposes unethical AI practices in advertising. They demand more transparency, see bias as a major issue, and feel manipulated by AI-driven ads. They strongly support bias detection mechanisms.

4.0.6 Cluster 2: Trusting & Less Concerned

1) Lowest privacy concern (Q7)

2) Believe ads are biased, but don't care much (Q14)

3) Low concerns about manipulation (Q11)

4) Moderate consent satisfaction (Q9)

5) Feel they have control over their ad experience (Q19)

Interpretation: This cluster is less worried about ethical concerns in AI advertising. They acknowledge bias in ads but don't feel strongly affected. They believe they have more control over ads, making them more accepting of AI-driven personalization.

```
[89]: for i in range(3):
          print(f"Sample respondents from Cluster {i}:")
          print(df_cluster[df_cluster['Cluster'] == i].head(3))
```

```
Sample respondents from Cluster 0:
    Q7: How concerned are you about the amount of personal data companies
collect to show personalized ads?  \
5                                                             4
6                                                             4
10                                                            4

    Q10: What level of transparency do you expect from companies that use AI for
personalized advertising?  \
5                                                             0
6                                                             0
10                                                            0

    Q14: Have you noticed that personalized ads are biased toward targeting
specific demographic groups?  \
5                                                             0
6                                                             0
10                                                            0

    Q15: Do you think incorporating bias detection mechanisms in AI systems can
improve fairness in personalized ads?  \
5                                                             2
6                                                             2
10                                                            1

    Q11: To what extent do you agree with the following statement: "Personalized
ads can manipulate my purchasing decisions by exploiting my preferences or
emotions."  \
5                                                             2
6                                                             2
10                                                            2

    Q9: When encountering personalized ads, do you feel you've given sufficient
consent for your data to be used?  \
5                                                             0
6                                                             0
10                                                            0
```

```
    Q19: Do you feel that you have adequate control over the types of
personalized ads you see?  \
5                                                              1
6                                                              0
10                                                             2


    Cluster
5         0
6         0
10        0
Sample respondents from Cluster 1:
   Q7: How concerned are you about the amount of personal data companies collect
to show personalized ads?  \
0                                                              3
1                                                              4
2                                                              3


    Q10: What level of transparency do you expect from companies that use AI for
personalized advertising?  \
0                                                              0
1                                                              0
2                                                              0


    Q14: Have you noticed that personalized ads are biased toward targeting
specific demographic groups?  \
0                                                              2
1                                                              2
2                                                              1


    Q15: Do you think incorporating bias detection mechanisms in AI systems can
improve fairness in personalized ads?  \
0                                                              2
1                                                              1
2                                                              2


    Q11: To what extent do you agree with the following statement: "Personalized
ads can manipulate my purchasing decisions by exploiting my preferences or
emotions."  \
0                                                              2
1                                                              2
2                                                              2


    Q9: When encountering personalized ads, do you feel you've given sufficient
consent for your data to be used?  \
0                                                              1
1                                                              0
2                                                              0
```

Q19: Do you feel that you have adequate control over the types of
personalized ads you see? \
0   2
1   1
2   2

    Cluster
0   1
1   1
2   1
Sample respondents from Cluster 2:
    Q7: How concerned are you about the amount of personal data companies
collect to show personalized ads? \
9   1
11   1
13   0

    Q10: What level of transparency do you expect from companies that use AI for
personalized advertising? \
9   2
11   0
13   1

    Q14: Have you noticed that personalized ads are biased toward targeting
specific demographic groups? \
9   2
11   2
13   2

    Q15: Do you think incorporating bias detection mechanisms in AI systems can
improve fairness in personalized ads? \
9   1
11   1
13   2

    Q11: To what extent do you agree with the following statement: "Personalized
ads can manipulate my purchasing decisions by exploiting my preferences or
emotions." \
9   1
11   2
13   4

    Q9: When encountering personalized ads, do you feel you've given sufficient
consent for your data to be used? \
9   1
11   1
13   0

```
     Q19: Do you feel that you have adequate control over the types of
personalized ads you see?  \
9                                                                    2
11                                                                   1
13                                                                   3

     Cluster
9          2
11         2
13         2
```

5 Association Rule Mining using Apriori and FP-Growth

```
[95]: # Step 1: Convert responses into a list of transactions (ensure all are strings)
      transactions = df.apply(lambda row: [str(item) for item in row.dropna()],␣
       ↪axis=1).tolist()
```

```
[96]: te = TransactionEncoder()
      te_ary = te.fit(transactions).transform(transactions)
      df_encoded = pd.DataFrame(te_ary, columns=te.columns_)
```

```
[97]: frequent_itemsets_apriori = apriori(df_encoded, min_support=0.2,␣
       ↪use_colnames=True)
```

```
[98]: rules_apriori = association_rules(frequent_itemsets_apriori, metric="lift",␣
       ↪min_threshold=1)
      rules_apriori = rules_apriori[(rules_apriori['confidence'] > 0.7) &␣
       ↪(rules_apriori['lift'] > 1.2)]
```

```
[99]: print("\n Top Apriori Rules:\n", rules_apriori.head(10))
```

```
 Top Apriori Rules:
                                         antecedents  \
18                                         (Student)
75                                               (4)
88                                               (5)
122                                          (Maybe)
126                                          (Maybe)
165                                          (Often)
196   (18-24, Full transparency about how data is co…
228                               (18-24, Sometimes)
235                                     (Student, 3)
280                                       (4, 18-24)

                           consequents  antecedent support  \
```

	antecedent	
18	(18-24)	0.419643
75	(Yes, they influence my decisions more)	0.428571
88	(Yes, frequently)	0.455357
122	(Not Sure)	0.276786
126	(Rarely)	0.276786
165	(Yes, they influence my decisions more)	0.419643
196	(3)	0.312500
228	(3)	0.276786
235	(18-24)	0.339286
280	(Yes, they influence my decisions more)	0.267857

	consequent support	support	confidence	lift	representativity
18	0.526786	0.348214	0.829787	1.575189	1.0
75	0.607143	0.312500	0.729167	1.200980	1.0
88	0.517857	0.321429	0.705882	1.363083	1.0
122	0.598214	0.205357	0.741935	1.240250	1.0
126	0.660714	0.232143	0.838710	1.269398	1.0
165	0.607143	0.312500	0.744681	1.226533	1.0
196	0.678571	0.258929	0.828571	1.221053	1.0
228	0.678571	0.241071	0.870968	1.283531	1.0
235	0.526786	0.276786	0.815789	1.548617	1.0
280	0.607143	0.214286	0.800000	1.317647	1.0

	leverage	conviction	zhangs_metric	jaccard	certainty	kulczynski
18	0.127152	2.780134	0.629191	0.582090	0.640305	0.745402
75	0.052296	1.450549	0.292857	0.432099	0.310606	0.621936
88	0.085619	1.639286	0.489071	0.493151	0.389978	0.663286
122	0.039780	1.556920	0.267848	0.306667	0.357706	0.542610
126	0.049267	2.103571	0.293447	0.329114	0.524618	0.595031
165	0.057717	1.538690	0.318242	0.437500	0.350097	0.629693
196	0.046875	1.875000	0.263323	0.353659	0.466667	0.605075
228	0.053253	2.491071	0.305441	0.337500	0.598566	0.613115
235	0.098055	2.568878	0.536181	0.469697	0.610725	0.670607
280	0.051658	1.964286	0.329268	0.324324	0.490909	0.576471

5.0.1 A. Student Users (Student → 18-24, Confidence = 82.98%)

Insight: A significant portion of students belong to the 18-24 age group.

Interpretation: Since this demographic is highly digitally active, their responses to personalized ads could shape advertising policies aimed at young adults.

5.0.2 B. Exposure to Personalized Ads → Decision Influence

Rule: (Ad Exposure Frequency = "4") → (Ads influence decisions more) (Confidence = 72.9%)

Rule: (Ad Exposure Frequency = "5") → (Yes, frequently) (Confidence = 70.5%)

Insight: The more frequently users see personalized ads, the more likely they are to acknowledge their influence.

Implication: This supports the argument that AI-driven ads can shape consumer choices. This raises concerns about manipulative advertising and highlights the need for transparency.

5.1 C. Uncertainty & Lack of Awareness

Rule: (Maybe) → (Not Sure) (Confidence = 74.1%)

Rule: (Maybe) → (Rarely) (Confidence = 83.8%)

Insight: Users who are uncertain ("Maybe") about AI-driven ad transparency tend to be unsure or disengaged when asked about its influence.

Implication: This suggests a lack of awareness or understanding about how AI-driven advertising affects them. Policies promoting clearer user education and consent mechanisms are necessary.

5.1.1 D. Transparency & Trust in AI Ads

Rule: (Full transparency about data collection) → (More engagement with ads) (Confidence = 82.8%)

Insight: Users who are aware of how AI collects and uses their data are more likely to engage.

Implication: Transparent AI advertising builds trust and can encourage ethical data-sharing behaviors.

5.1.2 E. Age and Decision Influence

Rule: (18-24, Sometimes engage with ads) → (More likely to be influenced) (Confidence = 87.0%)

Insight: The 18-24 age group is highly engaged with personalized ads, even if they don't actively seek them out.

Implication: This suggests that younger audiences are more susceptible to targeted advertising, highlighting a need for responsible AI-driven ad strategies.

```
[100]: frequent_itemsets_fp = fpgrowth(df_encoded, min_support=0.2, use_colnames=True)
```

```
[102]: rules_fp = association_rules(frequent_itemsets_fp, metric="lift",␣
       ↪min_threshold=1)
       rules_fp = rules_fp[(rules_fp['confidence'] > 0.7) & (rules_fp['lift'] > 1.2)]
```

```
[103]: print("\n Top FP-Growth Rules:\n", rules_fp.head(10))
```

```
 Top FP-Growth Rules:
                                          antecedents  \
291                          (Yes, Frequently, 3)
676                    (Occasionally, Not Sure, No)
692               (Occasionally, Not Sure, No, Yes)
696                    (Occasionally, Not Sure, No)
804  (Full transparency about how data is collected…
838  (3, Yes, Not Sure, Full transparency about how…
842  (3, Not Sure, Full transparency about how data…
```

```
924               (Occasionally, Yes, Frequently, Not Sure)
972   (Yes, they influence my decisions more, Not Su…
999                    (Yes, Frequently, Not Sure, 3)


                                consequents  antecedent support  \
291   (Yes, they influence my decisions more)             0.464286
676                                       (3)             0.321429
692                                       (3)             0.303571
696                                  (Yes, 3)             0.321429
804                            (Occasionally)             0.258929
838                            (Occasionally)             0.250000
842                       (Occasionally, Yes)             0.258929
924   (Yes, they influence my decisions more)             0.267857
972                                       (3)             0.267857
999   (Yes, they influence my decisions more)             0.276786


      consequent support   support   confidence       lift  representativity  \
291             0.607143  0.339286     0.730769   1.203620               1.0
676             0.678571  0.276786     0.861111   1.269006               1.0
692             0.678571  0.258929     0.852941   1.256966               1.0
696             0.660714  0.258929     0.805556   1.219219               1.0
804             0.669643  0.214286     0.827586   1.235862               1.0
838             0.669643  0.205357     0.821429   1.226667               1.0
842             0.642857  0.205357     0.793103   1.233716               1.0
924             0.607143  0.205357     0.766667   1.262745               1.0
972             0.678571  0.223214     0.833333   1.228070               1.0
999             0.607143  0.205357     0.741935   1.222011               1.0


      leverage  conviction  zhangs_metric   jaccard  certainty  kulczynski
291   0.057398    1.459184       0.315789  0.463415   0.314685    0.644796
676   0.058673    2.314286       0.312394  0.382716   0.567901    0.634503
692   0.052934    2.185714       0.293546  0.358025   0.542484    0.617260
696   0.046556    1.744898       0.264973  0.358025   0.426901    0.598724
804   0.040896    1.916071       0.257530  0.300000   0.478099    0.573793
838   0.037946    1.850000       0.246377  0.287500   0.459459    0.564048
842   0.038903    1.726190       0.255631  0.294872   0.420690    0.556274
924   0.042730    1.683673       0.284199  0.306667   0.406061    0.552451
972   0.041454    1.928571       0.253659  0.308642   0.481481    0.581140
999   0.037309    1.522321       0.251208  0.302632   0.343109    0.540085
```

5.1.3 A. High Exposure + Trust in AI Ads → Strongest Influence

Rule: (Ad Frequency = "3", "Yes, Frequently") → (More influence) (Confidence = 73.1%)

Insight: If users frequently see personalized ads and believe in AI-driven ads, they strongly acknowledge its influence.

Implication: AI-powered ads can reinforce purchasing decisions over time.

5.1.4 B. Uncertainty in AI Ads → Lower Trust & Engagement

Rule: (Occasionally see ads + Not Sure if AI Ads are Ethical) → (Disengagement) (Confidence = 85.2%)

Insight: Users who don't see AI-driven ads often and are unsure of its ethics tend to ignore personalized advertising.

Implication: This group may need educational campaigns about AI transparency and ethical advertising practices.

5.1.5 C. Transparency Matters Most for Ethical Ad Engagement

Rule: (Belief in Full Transparency) → (Occasionally Engages) (Confidence = 82.7%)

Insight: Users who know exactly how their data is collected are more comfortable engaging with ads.

Implication: AI-driven ads should clearly communicate data usage policies to gain user trust.