Here's a structured project plan with tasks for each analysis category. Each task includes objectives, suggested code, and potential insights to impress recruiters with a comprehensive analysis.

---

1. Household/Customer Analysis

Objective:

Understand customer purchasing patterns, average spending, and response to discounts. Identify loyal customers, high spenders, and discount-sensitive segments.

Tasks:

1. **Calculate Total and Average Spend per Household**

Code

```
household_spending = transactions.groupby('household_key')['sales_value'].agg(['sum', 'mean']).rename(columns={'sum': 'total_spend', 'mean': 'avg_spend'}).reset_index()

household_spending = household_spending.sort_values(by='total_spend', ascending=False)
```

   - **Insight**: Identify top-spending households and average spend, useful for targeting high-value customers.

2. **Analyze Discount Utilization by Household**

Code

```
household_discounts = transactions.groupby('household_key')['total_discount'].sum().reset_index()

household_discounts = household_discounts.sort_values(by='total_discount', ascending=False)
```

   - **Insight**: Determine households that benefit most from discounts, which can inform discount strategies.

3. **Determine Basket Size per Household**

Code

```
basket_size = transactions.groupby('household_key')['quantity'].sum().reset_index().rename(columns={'quantity': 'total_quantity'})
```

   - **Insight**: Find out if high-spending households also have larger basket sizes, helping with cross-sell and up-sell strategies.

---

2. Product Analysis

Objective:

Identify best-selling products, product categories that generate the most revenue, and product discount impact on sales.

Tasks:

1. **Identify Top-Selling Products by Quantity and Revenue**

Code

```
product_sales = transactions.groupby('product_id')['quantity', 'sales_value'].sum().reset_index()

product_sales = product_sales.sort_values(by='sales_value', ascending=False)

top_products = product_sales.head(10)
```

   - **Insight**: Recognize popular products and prioritize stock or promotions for these items.

2. **Analyze Discount Impact on Product Sales**

Code

```
product_discount_impact = transactions.groupby('product_id')['total_discount',
'sales_value'].sum().reset_index()

product_discount_impact['discount_to_sales_ratio'] = product_discount_impact['total_discount'] /
product_discount_impact['sales_value']
```

   - **Insight**: Determine which products are highly discount-driven, guiding discount allocation.

3. **Link with Product Details to Analyze Categories**

Code

```
product_details = pd.merge(product_sales, products[['product_id', 'department', 'brand']],
on='product_id', how='left')

category_sales =
product_details.groupby('department')['sales_value'].sum().reset_index().sort_values(by='sales_val
ue', ascending=False)
```

   - **Insight**: Identify top-performing product categories to inform assortment and stocking strategies.

---

3. Demographics Analysis

Objective:

Analyze spending patterns and discount sensitivity based on demographics like income, household size, and age group.

Tasks:

1. **Calculate Average Spend per Income Group**

Code

```
demographics_with_spend = pd.merge(demographics, household_spending, on='household_key', how='left')

income_spend = demographics_with_spend.groupby('income_desc')['total_spend'].mean().reset_index().sort_values(by='total_spend', ascending=False)
```

   - **Insight**: Target high-income groups with premium products and low-income groups with discount offers.

2. **Analyze Spend by Household Size**

Code

```
size_spend = demographics_with_spend.groupby('household_size_desc')['total_spend'].mean().reset_index().sort_values(by='total_spend', ascending=False)
```

   - **Insight**: Identify household sizes that spend the most, potentially useful for product bundling strategies.

3. **Analyze Spend by Age Group**

Code

```
age_spend = demographics_with_spend.groupby('age_desc')['total_spend'].mean().reset_index().sort_values(by='total_spend', ascending=False)
```

   - **Insight**: Discover spending patterns by age, which can inform age-targeted marketing.

---

4. Customer Loyalty and Retention Analysis

Objective:

Assess customer loyalty and develop strategies for retention by examining repeat purchases and changes in spending over time.

Tasks:

1. **Identify Repeat Customers**

Code

```
repeat_customers = transactions.groupby('household_key')['basket_id'].nunique().reset_index().rename(columns={'basket_id': 'total_baskets'})

repeat_customers = repeat_customers[repeat_customers['total_baskets'] > 1]
```

   - **Insight**: Quantify customer retention by finding households with multiple purchases.

2. **Analyze Monthly Spend Changes**

Code

```
transactions['month'] = transactions['day'] // 30  # Assuming 'day' is day count from the beginning of the dataset

monthly_spend = transactions.groupby(['household_key', 'month'])['sales_value'].sum().reset_index()

spend_changes = monthly_spend.pivot(index='household_key', columns='month', values='sales_value').fillna(0)
```

   - **Insight**: Track changes in household spending over time to identify trends and churn risks.

---

5. Discount Sensitivity Analysis

Objective:

Evaluate how discounts affect customer purchasing behavior across different demographics and product categories.

Tasks:

1. **Calculate Discount Utilization by Income Group**

Code

```
income_discount = demographics_with_spend.groupby('income_desc')['total_discount'].sum().reset_index().sort_values(by='total_discount', ascending=False)
```

- **Insight**: Identify which income groups are more responsive to discounts, helping to optimize promotional spending.

2. **Analyze Average Percentage Discount per Product Category**

Code

```
product_discounts = pd.merge(transactions, products[['product_id', 'department']], on='product_id', how='left')

category_discount = product_discounts.groupby('department')['percentage_discount'].mean().reset_index().sort_values(by='percentage_discount', ascending=False)
```

- **Insight**: Find categories with higher average discounts, potentially indicating high price sensitivity.

---

6. Visualizations and Reporting

Objective:

Create visualizations to communicate insights effectively. Visualize key metrics like top products, demographic spending patterns, and discount impacts.

Tasks:

1. **Bar Chart of Top Product Categories by Sales Value**

Code

```
import matplotlib.pyplot as plt

category_sales.plot(kind='bar', x='department', y='sales_value', title='Top Product Categories by Sales Value', color='skyblue')

plt.show()
```

2. **Box Plot of Sales Value Distribution by Household Size**

Code

```
demographics_with_spend.boxplot(column='total_spend', by='household_size_desc', vert=False, figsize=(10, 6), color='lightcoral')

plt.title('Sales Value Distribution by Household Size')

plt.suptitle('')

plt.show()
```

3. **Line Plot of Monthly Spend Changes for Top Households**

Code

```
spend_changes.T.plot(figsize=(12, 6), title='Monthly Spend Changes for Top Households')

plt.xlabel('Month')

plt.ylabel('Total Spend')

plt.show()
```