**Instacart Customer Profiling Project Report**

**Objective:** To support Instacart's marketing and business strategy teams by identifying meaningful customer profiles and behaviors, using customer, order, and product data.

**1. Data Import and Initial Exploration**

* Imported the following datasets: orders, products, orders\_products\_prior, and customers.
* Performed initial checks on data structure, missing values, and data types.
* Identified presence of null values (e.g., days\_since\_prior\_order) and minor duplications in some tables.

**2. Data Cleaning and Consistency Checks**

* Removed duplicate records from orders\_products\_prior.
* Dropped records in products with missing critical information such as product name or aisle ID.
* Accepted missing values in orders where applicable (e.g., NaN for users' first order).
* Cleaned and filtered customers dataset (if synthetic data was used, ensured completeness).
* Documented all issues and resolutions in the "Consistency Checks" tab.

**3. Data Wrangling**

* Mapped department\_id to descriptive department names for better interpretation.
* Converted columns to correct data types (e.g., integers, categories).
* Created a customer\_profile column using logic involving:
  + Age
  + Income
  + Number of dependents
  + Department behavior
  + Order time and day
* Ensured all personally identifiable data was removed or masked.

**4. Merging Datasets**

* Merged cleaned datasets in steps:
  1. orders + orders\_products\_prior → orders\_products\_combined
  2. Added products → orders\_products\_merged
  3. Added customers → orders\_products\_all
* Applied exclusion filter: max\_order < 5 to focus on active customers.
* Final merged dataset had approximately 31.9 million rows.

**5. Feature Engineering**

* Derived new columns:
  + customer\_profile
  + usage\_frequency
  + expenditure
* These were used to segment and analyze behavior across groups.

**6. Visual Analysis**

* Created visualizations to understand patterns:
  + Bar chart of customer profiles by region
  + Department preference by profile
  + Average expenditure by profile
* Included in the Visualizations tab and exported as PNGs.

**7. Insights and Recommendations**

* **Young Parents** were found mostly in the South, heavily shopping from the baby department.
* **Large Families** were common in North and East, favoring produce and household goods.
* **Single Adults** showed preference for gourmet items with higher spend per trip.
* **Retirees** had consistent shopping routines with moderate expenditure.

**Recommendations:**

* Tailor marketing campaigns by profile and region.
* Bundle products to match lifestyle (e.g., essentials for families).
* Promote gourmet or premium items to high-income singles.
* Time email offers and notifications to match ordering habits (e.g., retirees in the morning).

**Conclusion:** This project provided a foundational segmentation strategy using data-driven profiling. The results enable Instacart to design better-targeted promotions, refine regional offerings, and improve customer engagement through personalization.