Household Utility Billing Analysis Report

1. Introduction

The purpose of this project is to analyze household utility billing data to identify key trends, validate data integrity, and generate insights for process improvement. Using Google BigQuery, we performed SQL-based data validation, analysis, and aggregation to understand household payment behaviors.

2. Dataset Overview

The dataset contains information on household utility usage and payments. Below are the key columns:

- num_rooms: Number of rooms in the household
- num_people: Number of people living in the household
- housearea: Total area of the house
- is_tv: Whether the household owns a TV (1 = Yes, 0 = No)
- is_flat: Whether the house is a flat (1 = Yes, 0 = No)
- ave_monthly_income: Average monthly income of the household
- num_children: Number of children in the household
- is_urban: Whether the household is in an urban area (1 = Yes, 0 = No)
- amount_paid: Total amount paid for utilities

3. SQL Queries and Analysis

- Checking Total Records:

```
SELECT COUNT(*) AS total_records FROM utility_data.utility_bills;
```

- Checking for Missing Values:

```
SELECT * FROM utility_data.utility_bills WHERE num_rooms IS NULL OR amount_paid IS NULL;
```

- Average Monthly Income Analysis:

```
SELECT AVG(ave_monthly_income) AS avg_income FROM utility_data.utility_bills;
```

- Urban vs. Rural Household Payments:

```
SELECT is_urban, AVG(amount_paid) AS avg_amount_paid FROM utility_data.utility_bills GROUP BY
```

is_urban;

- Highest Paying Households:
  SELECT * FROM utility_data.utility_bills ORDER BY amount_paid DESC LIMIT 10;

4. Key Insights & Findings

- Urban households generally pay higher utility bills compared to rural households.

- Larger households tend to have higher utility expenses.

- Households with higher incomes also tend to have higher payments.

- Some records had missing values, which may require cleaning.

5. Recommendations & Next Steps

- Data Cleaning: Remove missing values and duplicates for accurate insights.

- Further Analysis: Investigate correlations between income, urban/rural status, and utility costs.

- Dashboard Creation: Visualize these findings using Google Data Studio or Mode Analytics.

- Predictive Modeling: Use Python or SQL ML to predict utility expenses based on household characteristics.

Prepared By: [Your Name]
Date: [Project Completion Date]