



# ***HOUSEHOLD DATASET***

November 24,2022

## **Team Members**

Mehmet Reşit ÇAĞAN 190709045

Ali DİKME 180702012.1

Murat Avzet ASAN 200709602

## **1 Purpose of Project**

Investigating specific expenditures of households in Philip-pines how relates on features of family using by “Household Dataset”.

## **2 Dataset Includes**

3 files, totally 81 columns and 15 strings in dataset

Family\_info.csv: [Click to redirect Kaggle](#)

House\_utilities.csv: [Click to redirect Kaggle](#)

Household expenses: [Click to redirect Kaggle](#)

### 3 Questions About Dataset

- 3.1 How many household head have of no job/with job (business indicator)?
- 3.2 Which households have income below the average and the number of vehicles above the average?
- 3.3 What is the rate of single family/extended family ?
- 3.4 What is the average number of family members ?
- 3.5 What is the rate of the members who under 5 ages/who between 5-17 ages ?
- 3.6 What is the mean of the family members number who has a job ?
- 3.7 What is the maximum education care expenditure ?
- 3.8 Which houses are priced above the average house price?
- 3.9 Which family has the least education expenses?
- 3.10 What is the mean of bread and cereals expenditure?

### 4 ER Diagram Link

[Click here to see ER diagram](#)

## 5 SQL Statements

### 5.1 What is the maximum education care expenditure?

```
select max(education)
from other_expenditure;
```

### 5.2 What is the mean of bread and cereals expenditure?

```
select avg(bread_and_cereals)
from food_expenditures;
```

### 5.3 Which households have income below the average and the number of vehicles above the average?

```
select id,household_income, vehicles.sum(count(*))
from financial_features,vehicles
where household_income < mean(household_income) in (
select id, sum(count(*))
from vehicles
where sum(count(*)) > mean(sum(count(*)))
);
```

#### 5.4 How many household head have no job/with job (business indicator)?

```
select count(head_job_or_business_indicator), head_job_or_business_indicator
from household_head
group by head_job_or_business_indicator;
```

#### 5.5 What is the average number of family members ?

```
select avg(number_of_members)
as FamilyMemberCountAverage
from family;
```

## 6 Loading Database With Values

We got our data from Kaggle. After the database setup, we will add data with the help of the front-end. We will take form-based input in the front-end part.

The trouble we may experience here will be the front end of the data operations. The data will be corrected and optimized in the front end. This will bring an extra workload.

Data will be added to the database with the help of an ORM(Entity Framework) via an API that works on the front-end and back end we have written.

## **7 Front-End User Interface**

We will implement front-end to get data from the user. We will use Blazor in order to write a MVC application. The technologies we will be using are HTML, CSS, JS, C# and .NET framework.

## **8 Database Platform and Challenging Issues**

We plan to use MySQL 8.0.31.CE on our home computer Apple M1 processor 8 GB of memory.

One of the issues we will be facing is to implement front-end user interface.

We did start splitting .csv files with R and will be finished after ER diagram will be approved. The issue here will be using regex to split compound variables to atomic. Because, all of team members are new to Regular Expression.