

For creating a data base, we created following four tables to connect the data:

For table "**households**":

1. we set a primary key called hh_id, and defined it as BIGINT since the households id are expected to be long integers.
2. We defined hh_race and is_latinx as INT because they will be integers indicating races.
3. We also defined hh_income and hh_size as INT because they contains integers indicating the income bracket and number of family members.
4. We used VARCHAR(5) to define hh_zip_code since zip code is 5-digit integer. Since we want to make sure that we don't drop the first digit in the zip code if it starts with a 0, we chose to consider zip code as a string of 5 characters.
5. We used VARCHAR(2) to define hh_state since it contains 2 character abbreviation of the state.
6. Finally, we used INT to define hh_residence_type because it contains integers indicating the residence type.

In the table **products**:

1. brand_at_prod_id represents the brand name of the each product belong to;
2. the department_at product_id includes the department information of the product;
3. the prod_id is every products ID number, since its contain a large number, therefore, we chose to use BIGINT for the data;
4. group_at_prod_id contains specific groups that product categorized as;
5. module shows more specific information about products categories;
6. and the amount shows the specific amount is;
7. the unit_at_prod_id indicates the units used for the amount measurement.

For table **trips**, it shows detail information of each trip for each household:

1. Household id is a large number so type "bigint" is selected;
2. TC_date is date with date, month and year when a trip happens so type "date" is selected.
3. TC_retailer_code is an integer so type "integer" is selected;
4. TC_retailer_code_store_code is code for each store so type "integer" is selected;
5. TC_retailer_code_store_zip3 is the zip code of each store at the retailer place so type "integer" is selected;
6. TC_total_spent is total consumption in each trip. The amount is rounded to 2 decimals so type "decimal" is selected;
7. TC_id is an integer and the primary key of the table so type "integer" is selected.

For table **purchases**, it shows the relationship of **products** and **trips** which is "many to many", because each household would buy different products in each trip, and one product would be included in different trips, so we have 6 attributes:

1. Two foreign keys, TC_id and prod_id, in this case we can have “one to many” relationship between table **Trips** and **Purchases**, and table **Products** and **Purchases**, respectively;
2. quantity_at_TC_prod_id, which is INT type;
3. total_price_paid_at_TC_prod_id, which is FLOAT type;
4. coupon_value_at_TC_prod_id, which is FLOAT type;
5. deal_flag_at_TC_prod_id, which is FLOAT type.

Here is the Scheme:

