

BUS211f1 Analyzing Big Data I

Fall 2019

Final Project: Analyzing Grocery Shopping Behavior

File 2 of 3

Part 1– Creating a data base

Create the database db_consumer_panel. The data for the 4 tables can be obtained in the files dta_at_hh.csv, dta_at_prod_id.csv, dta_at_TC_upc.csv, dta_at_TC.csv

- **Households:**
 - hh_id: long integer (PK)
 - hh_race: integer
 - 1: White Caucasian
 - 2: African American
 - 3: Asian
 - is_latinx: Integer indicating whether
 - 1 Yes
 - 2 No
 - hh_income: Integer indicating the income bracket.
 - 3: Under \$5k yearly income
 - 4: \$5k - \$7.9k
 - 6: \$8k - \$9.9k
 - 8: \$10k - \$11.9k
 - 10: \$12k – \$14.9k
 - 11: \$15k – \$19.9k
 - 13: \$20k – \$24.9k
 - 15: \$25k – \$29.9k
 - 16: \$30k – \$34.9k
 - 17: \$35k – \$39.9k
 - 18: \$40k – \$44.9k
 - 19: \$45k – \$49.9k
 - 21: \$50k – \$59.9k
 - 23: \$60k – \$69.9k
 - 26: \$70k – \$99.9k
 - 27: \$100.0k or more
 - hh_size: integer indicating the number of members composing the household.
 - 1: Single member
 - 2: Two members

- 3: Three members
 - ...
 - 9: Nine or more members
- hh_zip_code: 5 digits zipcode coded as integer
- hh_state: 2 character abbreviation of the state
- hh_residence_type:
 - 1: One family house
 - 2: One family house – condo
 - 3: Two family house
 - 4: Two family house - condo
 - 5: Three family house
 - 6: Three family house –condo
 - 7: Trailer
 - 8: Not reported.

| hh_id | hh_race | hh_is_latinx | hh_zip_code | hh_state | hh_income | hh_size | hh_residence_type |
|---------|---------|--------------|-------------|----------|-----------|---------|-------------------|
| 2000021 | 2 | 2 | 90746 | CA | 26 | 2 | 1 |
| 2000878 | 1 | 2 | 48912 | MI | 15 | 4 | 1 |

- **Products:**

- brand_at_prod_id
- department_at_prod_id
- prod_id (PK)
- group_at_prod_id
- module_at_prod_id
- amount_at_prod_id
- units_at_prod_id

| brand_at_prod_id | department_at_prod_id | prod_id | group_at_prod_id | module_at_prod_id | amount_at_prod_id | units_at_prod_id |
|------------------|-----------------------|--------------|------------------|-----------------------|-------------------|------------------|
| SINFUL COLORS | HEALTH & BEAUTY CARE | 1.000000e+16 | COSMETICS | COSMETICS-NAIL POLISH | 0.5 | OZ |
| SINFUL COLORS | HEALTH & BEAUTY CARE | 1.000000e+16 | COSMETICS | COSMETICS-NAIL POLISH | 0.5 | OZ |

- **Trips:**

- hh_id (FK)
- TC_date
- TC_retailer_code
- TC_retailer_code_store_code
- TC_retailer_code_store_zip3
- TC_total_spent:
- TC_id (PK)

| hh_id | TC_date | TC_retailer_code | TC_retailer_code_store_code | TC_retailer_code_store_zip3 | TC_total_spent | TC_id |
|---------|------------|------------------|-----------------------------|-----------------------------|----------------|-------|
| 2000000 | 2003-12-29 | 248 | 4125185 | 365.0 | 3.97 | 20576 |
| 2000000 | 2003-12-30 | 248 | 4125185 | 365.0 | 58.48 | 38829 |

- **Purchases:**

- TC_id (FK)
- quantity_at_TC_prod_id
- total_price_paid_at_TC_prod_id:
- coupon_value_at_TC_prod_id
- deal_flag_at_TC_prod_id
- prod_id (FK)

| TC_id | quantity_at_TC_prod_id | total_price_paid_at_TC_prod_id | coupon_value_at_TC_prod_id | deal_flag_at_TC_prod_id | prod_id |
|---------|------------------------|--------------------------------|----------------------------|-------------------------|--------------|
| 1030138 | 1 | 0.99 | 0.0 | 0 | 1.000000e+16 |
| 1030024 | 1 | 1.00 | 0.0 | 1 | 1.000000e+16 |