ALY6030 - Data Warehousing and SQL

CRN: 71719

Module 4 Assignment

By Karan Kaushal Daiya

Northeastern University: College of Professional Studies

Introduction:

For this task, we are utilizing information from Behavioral Risk Factor Surveillance System (BRFSS) to discover out alcohol use and its disorders among young people. BRFSS is a national framework of health-related phone surveys. It collects data of US inhabitants about their wellbeing conditions, impacts of alcohol and other behavioral conditions so that preventive measures can be taken. Here, we will use three datasets provided from BRFSS and basically focusing on Oklahoma state of US. In this assignment I have elaborated the task provided additionally we will develop a database plan for performing SQL queries.

**Task:** Exploring and gaining an understanding of data available from BRFSS data warehouse.

Table

Description automatically generated

Here we can see that in the sample size of 21, about 6.2 people are heavy drinker. This is high risk for young age group. Moreover, we can see from the dataset that higher income of break out has comparatively a smaller number of heavy drinkers among their sample size. The above given screenshot has the breakout category of age group from 18-24.

**Task:** Exploring survey responses (“Y”) that are relevant to adolescent alcohol consumption and some methods to identify them.

Liquor abuse can be figured out from the traits Sample\_size and Data\_value. Sample\_size is the number of individuals that replied “Yes”. Based on this we are able to calculate how many percentages of individuals from that group are overwhelming consumers. Moreover, the dataset contains diverse categories of individuals based on age-group, family salary, sexual orientation, race and ethnicity. Below query will identify groups of adolescents who may be at highest risk for alcohol abuse but there is no specific threshold that makes an age group considered to be high risk for alcohol consumption.

*CREATE SCHEMA `brfss` ;*

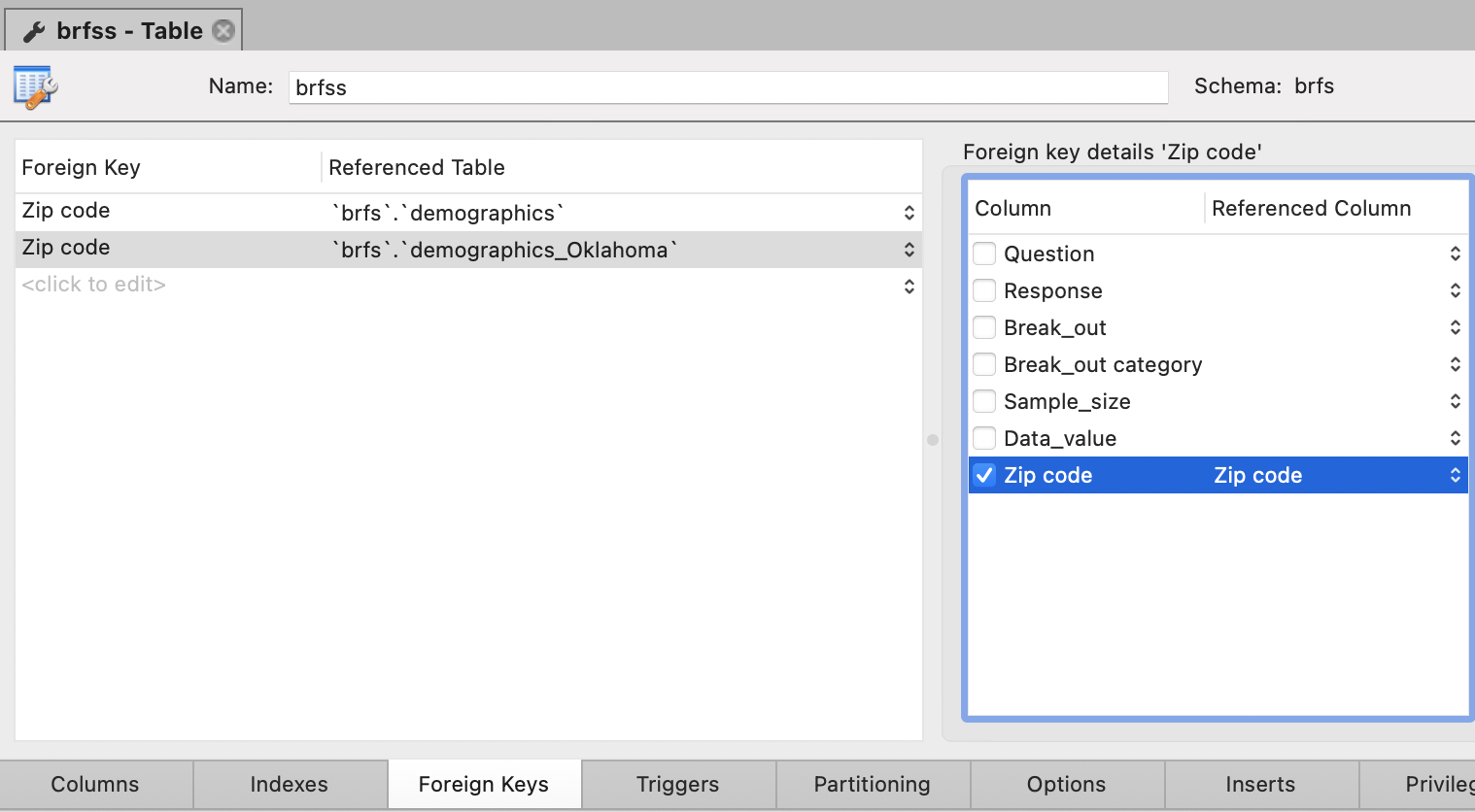
*SELECT AVG (Data\_value) AS No\_of\_drinkers FROM brfss\_ok WHERE Break\_out\_category= 'Age group' AND Break\_out='18-24' ;*

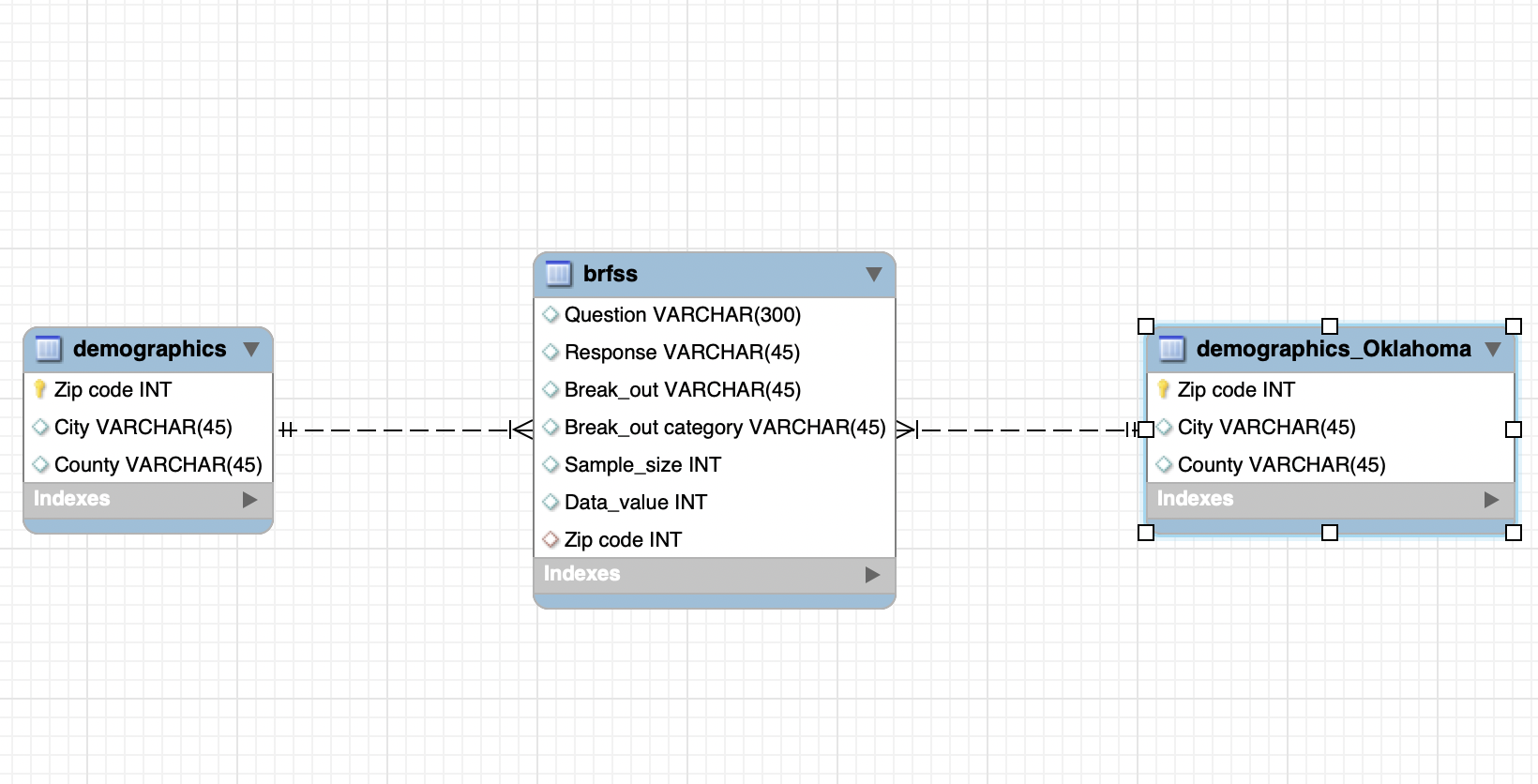
**ERD Design:**

Graphical user interface, text, application

Description automatically generated

We have created the tables and assigned primary keys, foreign keys and relationships. The ERD design for the database is as follows. This is in a general star schema.





Above shown is the screenshot of an ER Diagram of a star schema that we just created.

**Identifying areas of Oklahoma with highest and lowest number of respondents.**

Highest respondent,

SELECT MAX(Data\_value) AS highest\_respondant FROM brfss

INNER JOIN demographics\_Oklahoma ON brfss.Zipcode = demographics\_Oklahoma.Zipcode

WHERE Break\_outcategory='Age group' AND Break\_out= '18-24' ;

Lowest respondent,

SELECT MIN(Data\_value) AS highest\_respondant FROM brfss\_ok

INNER JOIN demographics\_Oklahoma ON brfss.Zipcode = demographics\_Oklahoma.Zipcode

WHERE Break\_outcategory='Age group' AND Break\_out= '18-24' ;

**Identifying areas of Oklahoma with highest and lowest number of respondents by city and county**

Highest respondent with a city,

SELECT MAX(Data\_value) AS highest\_respondant FROM brfss\_ok

INNER JOIN demographics\_Oklahoma ON brfss.Zipcode = demographics\_Oklahoma.Zipcode

WHERE Break\_outcategory='Age group' AND Break\_out= '18-24' AND City = 'CAMERON';

Lowest respondent with city,

SELECT MIN(Data\_value) AS highest\_respondant FROM brfss\_ok

INNER JOIN demographics\_Oklahoma ON brfss.Zipcode = demographics\_Oklahoma.Zipcode

WHERE Break\_outcategory='Age group' AND Break\_out= '18-24' AND City = 'CAMERON';

Highest respondent with a county,

SELECT MAX(Data\_value) AS highest\_respondant FROM brfss\_ok

INNER JOIN demographics\_Oklahoma ON brfss.Zipcode = demographics\_Oklahoma.Zipcode

WHERE Break\_outcategory='Age group' AND Break\_out= '18-24' AND County = LINCOLN;

Lowest respondent with a county,

SELECT MIN(Data\_value) AS highest\_respondant FROM brfss\_ok

INNER JOIN demographics\_Oklahoma ON brfss.Zipcode = demographics\_Oklahoma.Zipcode

WHERE Break\_outcategory='Age group' AND Break\_out= '18-24' AND County = LINCOLN;

Conclusion:

These sorts of datasets offer assistance to understand how liquor drinking is affecting diverse age individuals, distinctive classes of individuals, diverse race and ethnicity and so on. Numerous questions can be replied such as which ranges or age bunch, or households are exceedingly inclined to liquor mishandle, and which are not. Based on these information, legitimate activities can be taken to screen the circumstances and decrease the impacts of liquor on normal lives of individuals.