

D205

Assessment

Menachem Korn

A. Question of Interest

Which county has the least percent of customers with internet per state? Answering this question could help identify where to target marketing campaigns for internet providers.

A1. Meet the Data

The data needed to answer the question includes the customer and location tables from the churn database, alongside the services.csv file, to retrieve internet provider information.

The columns from the customer table needed are:

Customer_id, text primary key.

Location_id, Integer foreign key.

The columns from the location table needed are:

Location_id, Integer primary key.

State, text.

County, text.

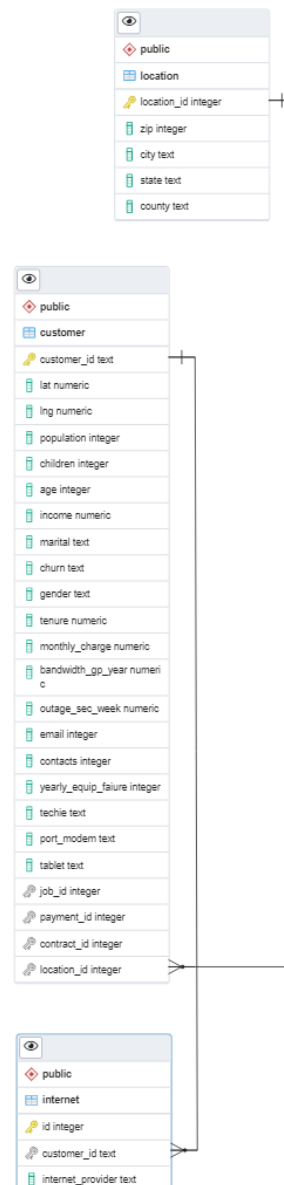
The columns used from the services.csv file are:

Customer_id, text foreign key.

internet_provider, text.

B. ERD

Included to the right is the ERD for
The tables used in the query.



B1. Creation of the Internet Table

Below is the SQL query used to create the table for the csv data:

```
CREATE TABLE internet (  
    id SERIAL PRIMARY KEY,  
    customer_id TEXT NOT NULL,  
    internet_provider TEXT NOT NULL,  
    CONSTRAINT fk_customer  
    FOREIGN KEY(customer_id)  
    REFERENCES  
    customer(customer_id),  
    CONSTRAINT customer_unique  
    UNIQUE(customer_id));
```

B2. Loading Data into the Internet Table

Below is the SQL query used to load data in the newly created table from B1. Due to the presence of extra columns in the csv file which were irrelevant to the query, a temporary table was created. All the data from the csv file was loaded into it, thereafter the relevant data was copied into the above table:

```
CREATE TEMPORARY TABLE t (  
    customer_id TEXT PRIMARY KEY,  
    InternetService TEXT,  
    Phone TEXT,  
    Multiple TEXT,  
    OnlineSecurity TEXT,  
    OnlineBackup TEXT,  
    DeviceProtection TEXT,  
    TechSupport TEXT );  
  
COPY T (customer_id,  
    InternetService,  
    Phone,  
    Multiple,  
    OnlineSecurity,  
    OnlineBackup,  
    DeviceProtection,  
    TechSupport)  
FROM 'C:\LabFiles\Services.csv'  
DELIMITER ','  
CSV HEADER;  
  
INSERT INTO internet (  
    customer_id,  
    internet_provider)  
SELECT customer_id, InternetService FROM t;
```

C. SQL Query

Attached is the query used to answer the question presented above; which county has the least percentage of customers with internet per state? For multiple counties within the same state having the same percentage, precedence was given to the county with the greatest number of records on file.

```
WITH percentage AS (  
  SELECT l.county,  
         l.state,  
         SUM(CASE  
           WHEN internet_provider = 'None'  
           THEN 0.0  
           ELSE 1.0  
         END) / COUNT(*) internet_percentage,  
         COUNT(*) cnt  
  FROM customer c  
  JOIN internet i  
    ON c.customer_id = i.customer_id  
  JOIN location l  
    ON c.location_id = l.location_id  
  GROUP BY county, state  
)  
ranked AS (  
  SELECT county,  
         state,  
         internet_percentage,  
         RANK() OVER(PARTITION BY state  
           ORDER BY internet_percentage, cnt DESC)  
  r from percentage)  
SELECT county, state, internet_percentage  
FROM ranked  
WHERE r=1  
ORDER BY state, county
```

C1. Output.

Attached is an output file of the above query named D205MenachemKorn.xls.

D. New Data

The data from the add on file must be updated right before the launch of a new marketing campaign, as well as daily throughout the campaign.

D1. Data Reload Intervals

The data should be updated before the launch of the campaign, to ascertain which clients changed their internet service provider due to marketing efforts. It would be counterintuitive to claim a customer signed up for internet due to the campaign, when possibly they signed up for service between the determination of marketing targets, and the public launch of the campaign.

So too, the data should be updated daily after public launch, to gain live insights into the success of the marketing efforts. The updated data may also be used for further A/B testing purposes, helping determine which approach garners the most subscriptions, adapting the strategy being used in future efforts. Through updating the data daily, we ensure the information gathered and insights observed will be the most up to date.

E. Panopto Videos

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=78f5ba9c-3a94-436b-8c3f-b0b1002c18d1>

E1. Panopto Video of Programs

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=1c9ed7ec-9779-46cb-838b-b0b1002e1a71>

F. Web Sources

No web sources or third-party code were used for the creation of this document or the completion of the assessment.

G. Web Sources

No sources or third-party code were used for the creation of this document or the completion of the assessment.