School of Engineering

Project Report

Jigyasa Dixit AU19B1003

Bachelor of Technology

(2022 – 2023)

Problem Statement:

An upcoming travel agency is having trouble scaling its business. They require a detailed analysis of their business using the data collected in the past, and current market trends and make future predictions and strategies.

Why does this problem matter?

With the increasing popularity of online tour websites, local agencies are suffering a great toll. Most people just search for tour plans and book them from these websites rather than getting a customised plan from a local one. This problem can be solved by increasing the visibility of these local agencies and giving more attractive offers and tour plans to the people so they would pay more attention to these companies.

Purpose:

1. Increasing the visibility of this agency among the locals.
2. Providing personalised tour plans to their customers.
3. Help the agency understand its audience and market trends better.
4. Ensure that the standard of the agency is subpar with the market standards.
5. Create an analysis based on their past tours and customers and help them make future plans.

I went to their office and gaining some insight into their work by asking questions related to their data and the future planning of their company. The questions are like:

- How many total trips go in a month?
- What is your busiest season?
- What is the most famous destination?
- Number of employees in a particular region.
- How many branches they currently have?
- What age group of people book the most trips.
- What are the seasons with the least trips?
- What are the platforms they are using right now for marketing?

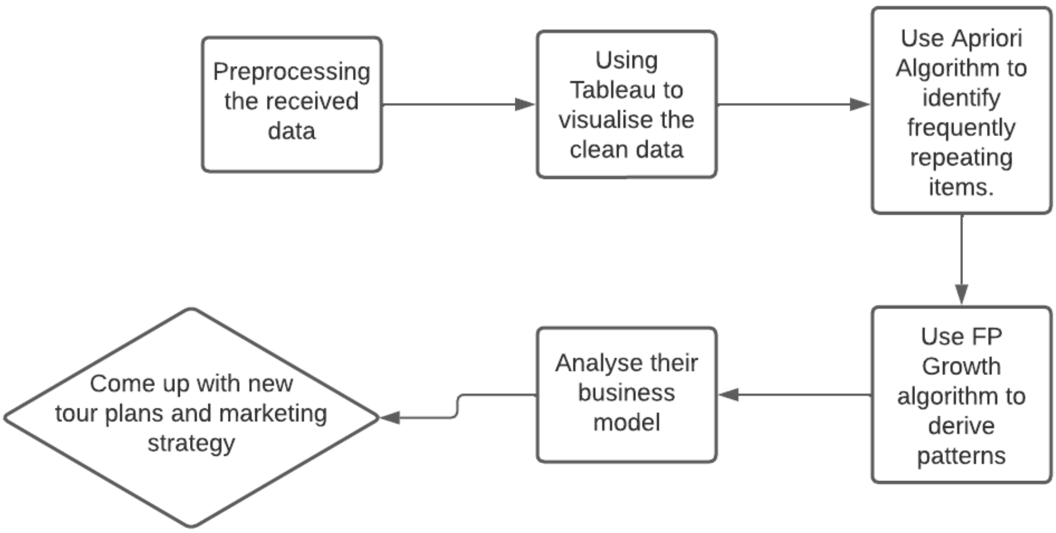
Based on the information acquired with the questions above, and hours of brainstorming, I came up with the following ideas to help them scale their business.

1. **Developing new marketing strategy:**
   We can help them reach new audience by developing a new marketing strategy for them.
   - Decide the platforms to use.
   - Design templates for the posts.
   - Help them create a brand identity with a colour palette.
   - Design a better logo.
   - Conduct in-depth exploration about the platforms' algorithm.
   - Understand and document all their past trips.
   - Find pictures from the past trips for beginning the posts.
   - Come up with new campaign ideas that can draw attention to their brand.
   - Collaborate with various travel influencers

- Paid advertisements
- Begin blogging and vlogging their travel experiences.

2. **Studying their past data**
   We can help them come up with new travel plans and tours by carefully studying and analysing their existing database.
   - Identify the peak travelling seasons
   - Identify the least travelled in season
   - Identify the age group of people booking the most tours
   - Come up with new tour plans for yet undiscovered places
   - Attract the younger audience by adding adventurous and fun activities in the itinerary.
   - Find out the reason for less tourists in a particular season
   - Identify the places that can be visited during those seasons
   - Plan more discounts in the least travelled in seasons

Flow Diagram:



Tools required to perform the aforementioned tasks:

1. Use Tableau for data visualization and analysis.
2. Use Apriori algorithm for more accurate analysis of the data.
3. Use FP Growth algorithm to understand the frequently repeating patterns in the dataset.
4. Google Survey to conduct more user research.
5. Content Marketing Tools to develop a marketing strategy

Comparative analysis between Apriori and FP-Growth algorithm

Apriori is a Join-Based algorithm and FP-Growth is a Tree-Based algorithm for frequent itemset mining or frequent pattern mining for market basket analysis. When it comes to marketing strategies it becomes very important to learn the behaviour of different customers regarding different products and services.

Apriori	**FP Growth**
Apriori generates frequent patterns by making the itemsets using pairings such as single item set, double itemset, and triple itemset.	FP Growth generates an FP-Tree for making frequent patterns.
Apriori uses candidate generation where frequent subsets are extended one item at a time.	FP-growth generates a conditional FP-Tree for every item in the data.
Since apriori scans the database in each of its steps it becomes time-consuming for data where the number of items is larger.	FP-tree requires only one scan of the database in its beginning steps so it consumes less time.
A converted version of the database is saved in the memory	Set of conditional FP-tree for every item is saved in the memory
It uses a breadth-first search	It uses a depth-first search.

Understanding from this comparative analysis, we conclude that using Apriori Algorithm would be more efficient in order to serve our purpose as it is less time consuming and is used to conduct market basket analysis. Compared to FP Growth in our case Apriori is a more appropriate algorithm to apply and hence more efficient in its results.

Applications of Apriori:

Some fields where Apriori is used:

1. **In Education Field:** Extracting association rules in data mining of admitted students through characteristics and specialties.
2. **In the Medical field:** For example Analysis of the patient's database.
3. **In Forestry:** Analysis of probability and intensity of forest fire with the forest fire data.
4. Apriori is used by many companies like Amazon in the **Recommender System** and by Google for the auto-complete feature.

Reference: [PDF] Research and Case Analysis of Apriori Algorithm Based on Mining Frequent Item-Sets (researchgate.net)

Analysis of existing travel agencies and services provided

MakeMyTrip:
The core value of MakeMyTrip is being guided by the customer's needs.
Products and Services provided:

- Flight Tickets
- Hotel Bookings
- Holiday Bookings
- Visa Services
- Bus Services
- Cab Services
- Train Services
- Home Stays
- MyBiz
- Air Bubble Flights

Yatra:
Their vision is to become the leading online travel company in India providing a "best in class" customer experience.
Products and services provided:

- Air Ticketing
- Hotels and packages
- Home Stays
- Rail Ticketing
- Bus Ticketing
- Cab Booking
- Cruise
- Cheap tickets to India
- Visa
- Hotels near me
- Mobile Apps

The services Gurukripa travels will provide that will distinguish them from the existing platforms:

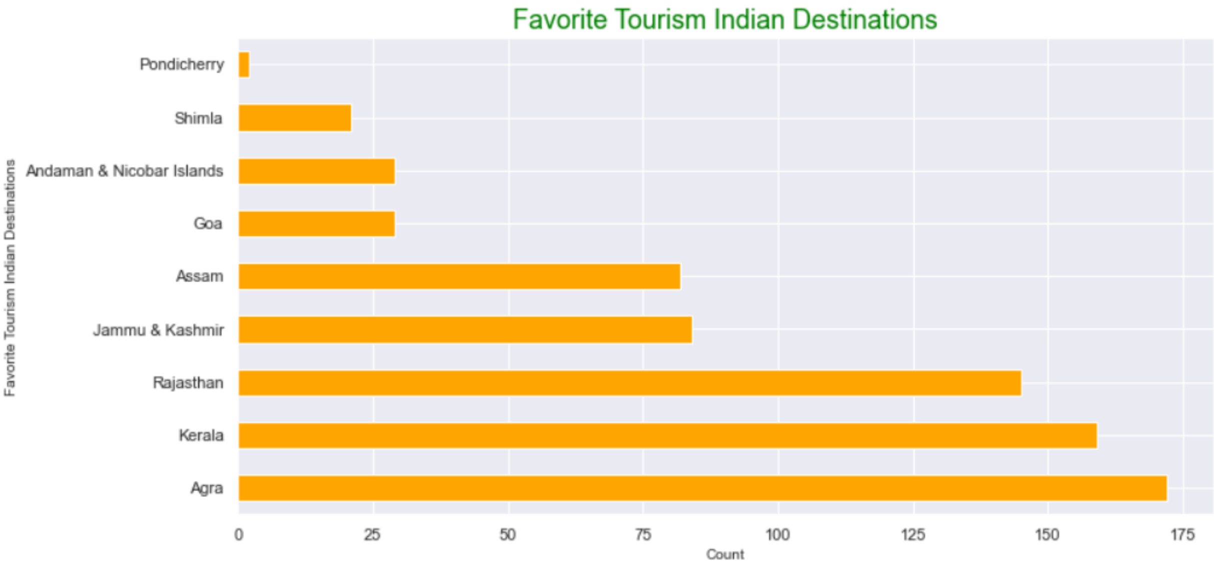
- Customised tour plan: As they will be conversing with the customers, either on call or face to face, they will be able to understand their precise requirements and make a better plan for them.

- Cabs at a lower rate: They will be providing their own vehicle for transportation at a lower rate along with a travel package.
- Annual trips: They will conduct various annual trips for different age groups based on their interests.
- Educational trips: They will attract more customers based on their interest and education they have received and make similar travel packages

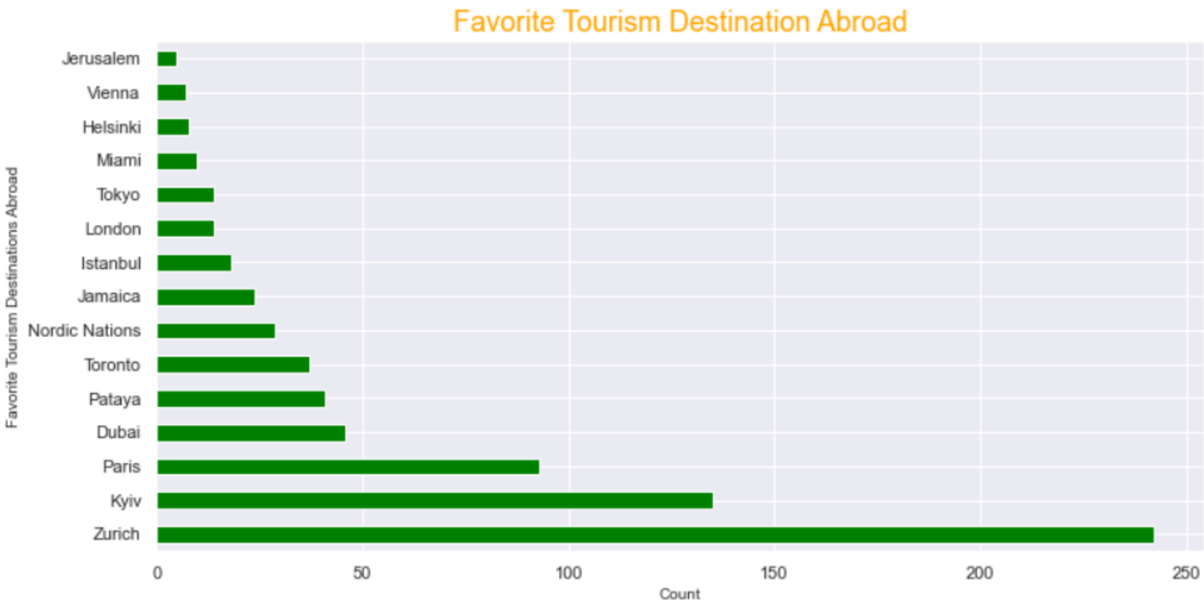
Basically, Gurukripa travels would be providing a much more personalised and comfortable experience as compared to the travel agencies online.

Analysis of their dataset:

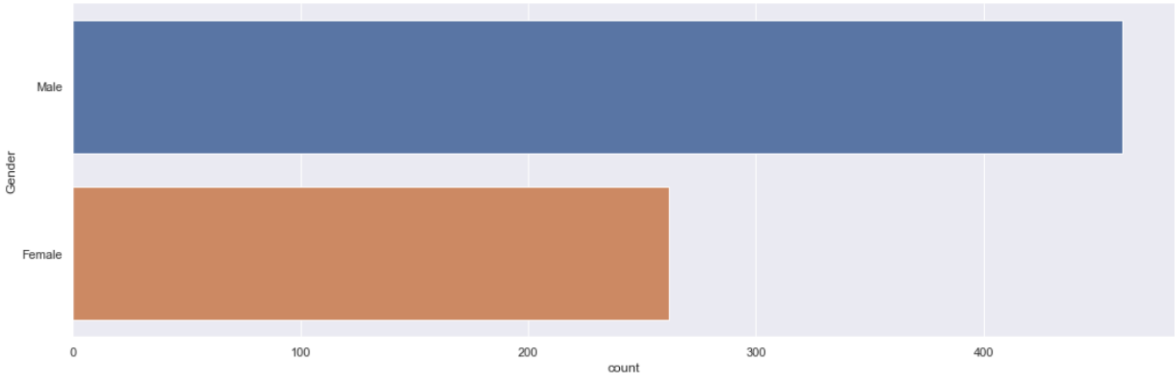
Favourite Tourism Destination in India:



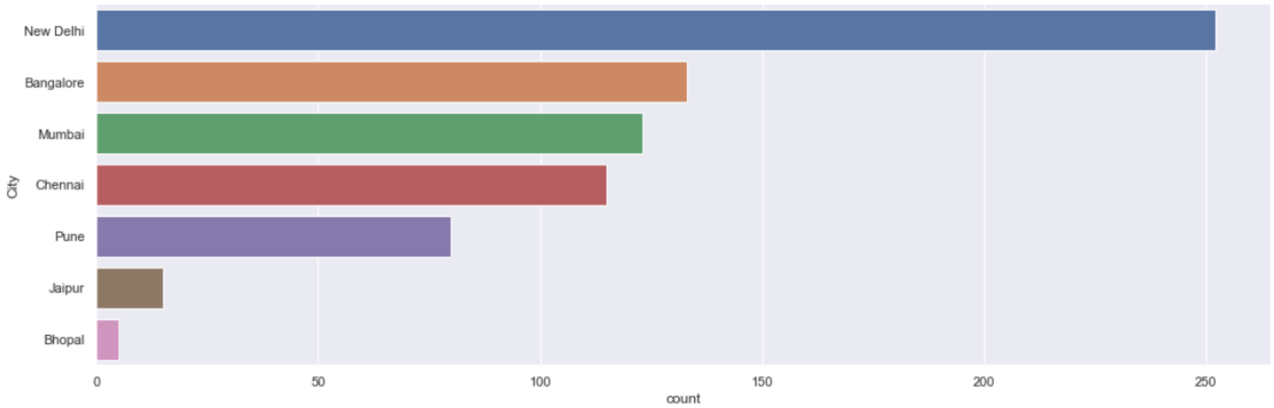
Favourite tourism destinations Abroad:



Tours segregated based on the genders:



Tours segregated based on city:



Applying Apriori Algorithm

Grouping by Age and favourite tourist destination:

```
Age_str = df.groupby(['Age', 'FavoriteTourismDest_india'])['FavoriteTourismDest_india'].count().reset_index(name ='Count')
Age_str
```

	Age	FavoriteTourismDest_india	Count
0	18	Agra	1
1	18	Rajasthan	1
2	19	Kerala	1
3	20	Jammu & Kashmir	1
4	21	Agra	4
...	...	...	...
119	38	Rajasthan	1
120	39	Jammu & Kashmir	1
121	39	Kerala	1
122	39	Rajasthan	1
123	40	Kerala	1

124 rows × 3 columns

making a matrice where m=age and n=FavoriteTourismDest_india and each row represents whether the item was in the transaction or not

```
# making a function which returns 0 or 1
# 0 means item was not in that transaction, 1 means item present in that transaction

def encode(x):
    if x<=0:
        return 0
    if x>=1:
        return 1

# applying the function to the dataset

my_basket_sets = my_basket.applymap(encode)
my_basket_sets.head()
```

FavoriteTourismDest_india	Agra	Andaman & Nicobar Islands	Assam	Goa	Jammu & Kashmir	Kerala	Pondicherry	Rajasthan	Shimla
Age									
18	1	0	0	0	0	0	0	1	0
19	0	0	0	0	0	1	0	0	0
20	0	0	0	0	1	0	0	0	0
21	1	0	1	0	1	1	0	1	0
22	1	0	1	0	1	1	1	1	0

Using the 'apriori algorithm' with min_support=0.01 (1% of 9465)

It means the item should be present in atleast 94 transaction out of 9465 transactions only when we considered that item in frequent itemset

```
frequent_items = apriori(my_basket_sets, min_support = 0.01,use_colnames = True)
frequent_items
```

Out[29]:

	support	itemsets
0	0.782609	(Agra)
1	0.391304	(Andaman & Nicobar Islands)
2	0.695652	(Assam)
3	0.565217	(Goa)
4	0.782609	(Jammu & Kashmir)
...	...	...
506	0.043478	(Shimla, Rajasthan, Kerala, Jammu & Kashmir, A...
507	0.043478	(Shimla, Rajasthan, Goa, Kerala, Jammu & Kashm...
508	0.043478	(Shimla, Rajasthan, Goa, Kerala, Jammu & Kashm...
509	0.043478	(Shimla, Rajasthan, Goa, Kerala, Jammu & Kashm...
510	0.043478	(Shimla, Rajasthan, Goa, Kerala, Jammu & Kashm...

511 rows × 2 columns

Association Rules and Frequent Item sets

The Apriori algorithm calculates rules that express probabilistic relationships between items in frequent item sets, for example, a rule derived from frequent item sets containing A, B, and C might state that if A and B are included in a transaction, then C is likely to also be included.

An association rule states that an item or group of items implies the presence of another item with some probability. Unlike decision tree rules, which predict a target, association rules simply express correlation.

Antecedent and Consequent

The IF component of an association rule is known as the **antecedent**. The THEN component is known as the **consequent**. The antecedent and the consequent are disjoint; they have no items in common.

Oracle Data Mining supports association rules that have one or more items in the antecedent and a single item in the consequent.

Confidence

Rules have an associated, which is the conditional probability that the consequent will occur given the occurrence of the antecedent. The minimum confidence for rules can be specified by the user.

Reference: https://docs.oracle.com/cd/E24693_01/datamine.11203/e16808/algo_apriori.htm

Now making the rules from frequent itemset generated above:

```
rules = association_rules(frequent_items, metric = "lift", min_threshold = 1)
rules.sort_values('confidence', ascending = False, inplace = True)
rules
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Andaman & Nicobar Islands)	(Agra)	0.391304	0.782609	0.391304	1.00	1.277778	0.085066	inf
16192	(Shimla, Rajasthan, Agra, Assam, Pondicherry)	(Goa, Andaman & Nicobar Islands, Jammu & Kashmir)	0.043478	0.391304	0.043478	1.00	2.555556	0.026465	inf
7907	(Shimla, Rajasthan, Agra, Pondicherry, Andaman...	(Jammu & Kashmir)	0.043478	0.782609	0.043478	1.00	1.277778	0.009452	inf
7908	(Shimla, Jammu & Kashmir, Agra, Pondicherry, A...	(Rajasthan)	0.043478	0.826087	0.043478	1.00	1.210526	0.007561	inf
7909	(Rajasthan, Jammu & Kashmir, Agra, Pondicherry...	(Shimla)	0.043478	0.391304	0.043478	1.00	2.555556	0.026465	inf
...	...	...	...	...	...	...	...	...	...
14996	(Kerala)	(Shimla, Rajasthan, Goa, Assam, Pondicherry, A...	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
14491	(Kerala)	(Rajasthan, Goa, Jammu & Kashmir, Assam, Pondi...	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
1277	(Kerala)	(Agra, Pondicherry, Shimla)	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
15374	(Kerala)	(Shimla, Rajasthan, Goa, Jammu & Kashmir, Assa...	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
1848	(Kerala)	(Goa, Assam, Pondicherry)	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865

18174 rows × 9 columns

Arranging the data from highest to lowest with respect to 'confidence':

```
rules.sort_values('confidence', ascending=False)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Andaman & Nicobar Islands)	(Agra)	0.391304	0.782609	0.391304	1.00	1.277778	0.085066	inf
10811	(Jammu & Kashmir, Shimla, Pondicherry)	(Rajasthan, Kerala, Assam)	0.043478	0.695652	0.043478	1.00	1.437500	0.013233	inf
15126	(Shimla, Rajasthan, Goa, Kerala, Jammu & Kashm...	(Andaman & Nicobar Islands)	0.043478	0.391304	0.043478	1.00	2.555556	0.026465	inf
15102	(Shimla, Pondicherry)	(Rajasthan, Kerala, Jammu & Kashmir, Assam, An...	0.043478	0.391304	0.043478	1.00	2.555556	0.026465	inf
10806	(Rajasthan, Shimla, Pondicherry)	(Jammu & Kashmir, Kerala, Assam)	0.043478	0.695652	0.043478	1.00	1.437500	0.013233	inf
...	...	...	...	...	...	...	...	...	...
4000	(Kerala)	(Rajasthan, Goa, Agra, Pondicherry)	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
5303	(Kerala)	(Goa, Assam, Jammu & Kashmir, Pondicherry)	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
9879	(Kerala)	(Rajasthan, Jammu & Kashmir, Assam, Pondicherr...	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
17151	(Kerala)	(Shimla, Rajasthan, Goa, Jammu & Kashmir, Agra...	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865
1848	(Kerala)	(Goa, Assam, Pondicherry)	0.869565	0.043478	0.043478	0.05	1.150000	0.005671	1.006865

Explaining the output:

Deriving patterns from the available data we can say that if a person has travelled to Andaman & Nicobar Islands, then based on customer history there is a high possibility that they will be interested in travelling to Agra as well.

Algorithm applied after grouping by Gender and Favourite tourism destination in India

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Andaman & Nicobar Islands)	(Agra)	1.0	1.0	1.0	1.0	1.0	0.0	inf
1640	(Rajasthan, Shimla, Pondicherry)	(Andaman & Nicobar Islands)	0.5	1.0	0.5	1.0	1.0	0.0	inf
1631	(Andaman & Nicobar Islands, Kerala)	(Rajasthan, Shimla)	1.0	1.0	1.0	1.0	1.0	0.0	inf
1632	(Andaman & Nicobar Islands, Shimla)	(Rajasthan, Kerala)	1.0	1.0	1.0	1.0	1.0	0.0	inf
1633	(Kerala, Shimla)	(Rajasthan, Andaman & Nicobar Islands)	1.0	1.0	1.0	1.0	1.0	0.0	inf
...	...	...	...	...	...	...	...	...	...
2808	(Kerala)	(Rajasthan, Andaman & Nicobar Islands, Agra, P...	1.0	0.5	0.5	0.5	1.0	0.0	1.0
2879	(Rajasthan, Andaman & Nicobar Islands, Shimla)	(Agra, Pondicherry)	1.0	0.5	0.5	0.5	1.0	0.0	1.0
2831	(Agra, Kerala)	(Andaman & Nicobar Islands, Shimla, Pondicherry)	1.0	0.5	0.5	0.5	1.0	0.0	1.0
2877	(Rajasthan, Agra, Shimla)	(Andaman & Nicobar Islands, Pondicherry)	1.0	0.5	0.5	0.5	1.0	0.0	1.0
10167	(Andaman & Nicobar Islands)	(Shimla, Rajasthan, Goa, Kerala, Jammu & Kashm...	1.0	0.5	0.5	0.5	1.0	0.0	1.0

10168 rows × 9 columns