AIRBNB ANALYTICS CASE STUDY

BY SURABHI GOYAL

Agenda

- Introduction
- Overview
- Objectives
- Dataset Details
- Data Analysis In MySQL
- Business Insights

INTRODUCING

SURABHI GOYAL

Result-oriented Data Analyst with a proven ability to extract meaningful insights from complex data. Skilled in SQL, Python, and data visualization tools, I leverage statistical analysis and data mining techniques to drive business decisions.

OVERVIEW

The project involves leveraging extensive databases within a dynamic e-commerce company to extract actionable insights. The objective is to inform key departments-such as marketing and supply chain-on strategic decisions that optimize operations, improve customer satisfaction, and enhance sales performance. This initiative aims to utilize data analysis to drive forward the company's business strategies effectively.

OBJECTIVES

- User Behavior and Engagement: Identify key user segments based on demographics, behavior, and preferences.
- Geographic Insights: Analyze geographic distribution of users to identify key markets.
- Sales Optimization: Analyzing sales data to identify trends, opportunities, and areas for improvement.
- Platform Performance: Track key performance indicators (KPIs) such as booking rates, conversion rates, and revenue.

DATASET DETAILS

- **Users Table:** This table contains comprehensive user data for a travel or booking platform, encompassing personal details, account information, marketing interactions, device usage, and travel preferences.
- **Sessions Table:** Analyze geographic distribution of users to identify key markets.
- **Countries Table:** This table provides insights into destination countries, including their geographical locations, distances from users, area sizes, primary languages spoken, and a measure of language similarity.

DATA ANALYSIS IN MYSQL

Problem Statement:
Identify the top 5 most active users who have spent more than 10,000 seconds on at least one session.
'Most active' is defined as having the highest number of sessions.
This will help you analyze user session data to find a potential correlation between session duration and user activity.

```
select user_id from (select user_id, count(*) as session_count, max(secs_elapsed) as high_value from sessions_data
group by user_id having high_value > 10000 order by session_count desc) temp limit 5;
```

user_id
l4761qsja1
ep4zku17of
jpd04cjpz7
x1n1xxlfzj
tjjpohec99

Problem Statement:
determine the most frequently used signup method for each Gender category, considering only users who have made a booking (as indicated by a non-null value in the Date_first_booking column).
This exploration will help us understand if certain demographic factors are associated with specific signup preferences among users who follow through with bookings.

```
select gender, signup_method, count(*) from users where Date_first_booking is not null and country_destination <> 'NDF'
 group by 1,2;
```

gender	signup_method	count(*)
MALE	facebook	130
FEMALE	basic	252
FEMALE	facebook	137
MALE	basic	185
	basic	321
	facebook	8
MALE	google	1
OTHER	basic	1

Problem Statement:
Determine the average age of users by destination country, considering only those with a booking and available age data.
Sort the results from the youngest to the oldest users.
This will help you understand the destination country preferences for different age of users.

```
select country_destination, avg(age) as average_age from users where age is not null and
date_first_booking is not null and country_destination <> 'NDF' group by 1
order by 2;
```

country_destination	average_age
AU	30.0000
NL	30.5000
DE	33.5000
FR	34.4737
CA	35.8182
ES	37.6111
other	38.1802
US	40.0163
PT	105.0000
GB	151.4706
IT	200.8333

Problem Statement:
Write a query to find users with fewer than 5 sessions who made a booking to the destination "US".
Sort the results by the number of sessions in descending order.

```sql
select s.user_id , count(s.user_id) as session_count from sessions_data s join users u on s.user_id = u.id
where u.country_destination = 'US' group by 1
having count(s.user_id) < 5 order by 2 desc;
```

user_id	session_count
rif93dskj5	4
m4ob9byg62	4
ziz0zrvwco	4
evd5ad8ydd	4
26xz1gkd0a	3
mwvd6ufzyb	3
i0602o51ji	3
qsmrdwjtsx	2
xq020ca3aq	2
6fjv7jzvm3	2
3ajimh98kd	2

Problem Statement:
Write a SQL query to identify the top 5 most common actions performed by users who made a booking (i.e., country_destination is not 'NDF') and the devices they use for these actions.
This information can help in optimizing the user experience and tailoring the interface to common user behaviors.

```
select s.action, s.device_type, count(s.action)  as action_count from sessions_data s join users u
on u.id = s.user_id where u.country_destination <> 'NDF' and
u.date_first_booking is not null group by 1,2 order by 3 desc limit 5;
```

action	device_type	action_count
show	Mac Desktop	2297
show	Windows Desktop	1257
personalize	Mac Desktop	1240
show	iPhone	1219
index	Mac Desktop	1031

Problem Statement:
Write a SQL query to find the most frequent combinations of two actions (performed by the same user on Windows Desktop devices) where the most time is spent, for users who have made a booking (i.e., country_destination is not 'NDF'). Consider the top 10 combinations from the resulting table which will be considered as most frequent.

```
select s1.action as first_action, s2.action as second_action , count(*) as action_pair_count ,
sum(s1.secs_elapsed + s2.secs_elapsed) as total_time_spent from sessions_data s1 join sessions_data s2
on s1.user_id = s2.user_id and s1.action <> s2.action join users u on s1.user_id = u.id
where s1.device_type = "Windows desktop" and s2.device_type = "Windows desktop" and
u.country_destination <> 'NDF' group by s1.action , s2.action order by 4 desc limit 10;
```

first_action	second_action	action_pair_count	total_time_spent
show	index	23355	1153786637
index	show	23355	1153786637
index	search_results	10650	546803044
search_results	index	10650	546803044
search_results	show	16883	464799772
show	search_results	16883	464799772
personalize	show	24867	448969000
show	personalize	24867	448969000
ajax_refresh_subtotal	show	20490	418433689
show	ajax_refresh_subtotal	20490	418433689

Problem Statement:
To understand which affiliate channels are most effective, analyze the number of bookings made through each first affiliate channel and calculate their conversion rates.
Write an SQL query to find the number of bookings and the conversion rate for each first affiliate channel. Consider a booking as made if country_destination is not 'NDF'.

```sql
SELECT first_affiliate_tracked AS affiliate_channel, COUNT(*) AS total_users, SUM(country_destination <> 'NDF') AS bookings,
(SUM(country_destination <> 'NDF') / COUNT(*)) * 100 AS conversion_rate
FROM users GROUP BY first_affiliate_tracked
ORDER BY conversion_rate DESC;
```

affiliate_channel	total_users	bookings	conversion_rate
untracked	1288	586	45.4969
linked	555	229	41.2613
tracked-other	73	29	39.7260
omg	500	167	33.4000
	64	19	29.6875
product	17	5	29.4118
marketing	3	0	0.0000

Problem Statement:
To further understand the effectiveness of different affiliate providers and signup methods, determine the conversion rate for each combination.
Write a SQL query to calculate the conversion rate for each combination of affiliate provider and signup method. Consider a booking as made
if country_destination is not 'NDF'.

```sql
select affiliate_provider, signup_method, count(*) as total_users,
sum(country_destination <> 'NDF') as bookings , (sum(country_destination <> 'NDF')/count(*))*100
as conversion_rate from users group by 1,2 order by 5 desc;
```

affiliate_provider	signup_method	total_users	bookings	conversion_rate
padmapper	facebook	5	4	80.0000
facebook-open-graph	facebook	6	3	50.0000
meetup	basic	2	1	50.0000
facebook-open-graph	basic	2	1	50.0000
craigslist	facebook	21	10	47.6190
direct	basic	1207	552	45.7332
craigslist	basic	29	12	41.3793
direct	facebook	412	167	40.5340
facebook	basic	10	4	40.0000
other	basic	59	23	38.9831
google	facebook	144	55	38.1944

Problem Statement:
Write a SQL query to assess the effectiveness of different marketing channels by calculating the conversion rate for each affiliate channel.
Consider a booking as made if country_destination is not 'NDF'.

```
select affiliate_channel, count(*) as total_users, sum(country_destination <> 'NDF') as bookings ,
(sum(country_destination <> 'NDF')/count(*))*100
as conversion_rate from users group by 1 order by 4 desc;
```

affiliate_channel	total_users	bookings	conversion_rate
seo	105	47	44.7619
other	117	52	4 44.4444
direct	1627	720	44.2532
sem-brand	285	107	37.5439
api	98	33	33.6735
remarketing	18	6	33.3333
sem-non-brand	218	68	31.1927
content	32	2	6.2500

```
-- SEO is the affiliate channel which has the highest conversion rate.
```

BUSINESS INSIGHTS

• 'US’ and 'France' are those destinations where mostly users visited so these countries to be focused on as part of the marketing strategies.

• 'iPhones' are the most common device for users overall, but for users who spend the most time on the platform, 'Mac desktops' dominate. This suggests 'Mac' users might be more engaged in the planning and browsing process.

• Analyzing sessions by user helps identify "power users" with high session counts and long session durations. Exploring their behavior can provide valuable insights into user journeys and areas for improvement.

• Most frequent signup method for female users with bookings is 'Basic' and for male users, it is 'Facebook' which can be a good target for marketing efforts.

• Age group of '30-40' prefers their destination countries are ‘Australia’, 'Netherlands', 'Germany' and '40-more' prefers 'US', 'Portugal’.

• Total clicks made by organic users are '8469' which helps to assess its overall activity and potential impact.

• 'Show' and 'Index' are the actions type in which most time was taken by the user sequentially which help optimize the user journey by focusing on significant action pairs on Windows Desktop.

• 'Untracked' and 'Linked' are the most affiliate channels for user acquisition (conversion rate is the highest) which helps us to inform marketing budget allocation and partnerships.

• Combination of affiliate provider and signup method is ‘Pad-mapper' and 'Facebook' which gives the highest conversion rate and hence it is the best to tailor user acquisition strategies for specific demographics and marketing channels.

Recommendations based on the insights:

Cater the interface to browsing and planning activities, potentially on larger screens.

• Analyze and understand power user behavior to identify potential pain points and areas for improvement.

• Use location and user demographics to tailor marketing campaigns for specific regions and user groups.

• Promote the most effective signup methods for different user segments based on demographics or referral sources.

• Implement strategies to encourage organic user engagement and conversion.

•Analyze actions with low average time spent and consider content or interface improvements.

• Allocate marketing budget and resources to affiliate channels with the highest conversion rates.

• Test different signup methods for each affiliate provider to optimize conversion rates.

THANK YOU!

+91-8958936840

surabhigoyal198@gmail.com

https://github.com/SurabhiGoyal-Coder

https://www.linkedin.com/in/surabhi-goyal-36154a207