

Tasks

Task 1: Analysis

1a- Bookings

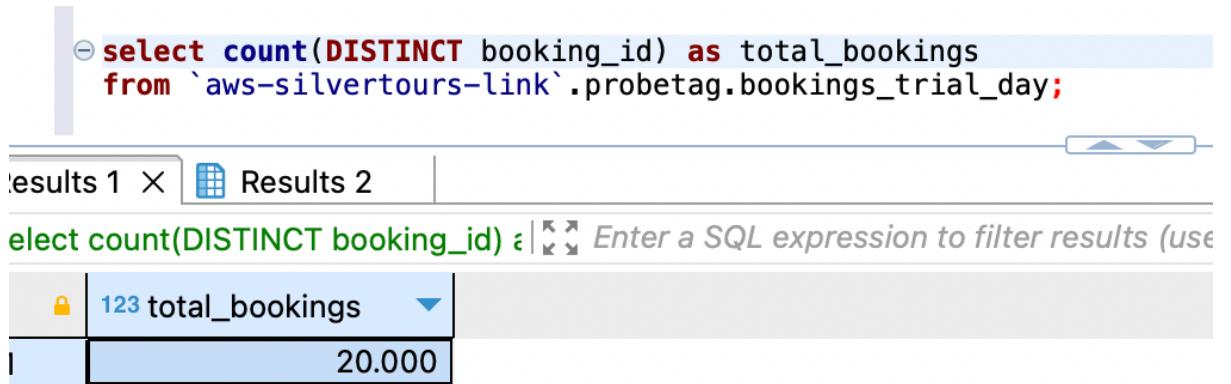
Please have a look at the bookings table to understand the data inside. Please make the difference between cancelled and non-cancelled bookings.

1. Do you notice anything relevant in the data?

There is a flag “booking_is_cancelled” in the table to differentiate between cancelled and non-cancelled Bookings. True representing a cancelled booking and False representing a non-cancelled booking.

2. How many bookings have been created?

There are total 20,000 bookings in the Bookings Table.



The screenshot shows a SQL query result table. The query is:

```
select count(DISTINCT booking_id) as total_bookings  
from `aws-silvertours-link`.probetag.bookings_trial_day;
```

The results table has two columns: 'total_bookings' and a lock icon. The value '20.000' is displayed in the 'total_bookings' column.

total_bookings
20.000

3. How many departures?

Assuming number of departures is the number of non-cancelled bookings, we have total 15,104 number of departures.

```
④ select count(DISTINCT booking_id) as total_departures  
from `aws-silvertours-link`.probetag.bookings_trial_day  
WHERE booking_is_cancelled is FALSE ;
```

Results 1 × Results 2 | Enter a SQL expression to filter results

	123 total_departures
	15.104

4. How many customers?

There are total 18,721 unique customers.

```
④ select count(DISTINCT customer_id) as total_customers  
from `aws-silvertours-link`.probetag.bookings_trial_day ;
```

Results 1 × Results 2 | Enter a SQL expression to filter results

	123 total_customers
1	18.721

5. For each booking, what is the average price per day based on the trip length?

I have devised two queries to serve the same purpose.

- One regular,

```
④ SELECT booking_id,  
booking_price / (DATE_DIFF(DATE(booking_dest_date), DATE(booking_dep_date), DAY) + 1) AS average_price_per_day  
FROM `aws-silvertours-link`.probetag.bookings_trial_day ;
```

- One with Common table expression,

```

-- With Common table Expression
WITH price_per_booking AS (
    SELECT
        booking_id, booking_price,
        DATE_DIFF(DATE(booking_dest_date), DATE(booking_dep_date), DAY) + 1 AS booking_duration
    FROM
        `aws-silvertours-link`.probetag.bookings_trial_day
)
SELECT
    booking_id,
    booking_price / booking_duration AS average_price_per_day
FROM
    price_per_booking;

```

Results 1 ×

WITH price_per_booking AS (SELECT Enter a SQL expression to filter results (use Ctrl+Space)

Grid	ABC booking_id	123 average_price_per_day
1	3498b368819e9dd3dbcb2774e3228d9a6df0d733	73,9499969482
2	08efe051fbceb577fe579b2cf005eed8d53fe0e0	32,9187507629
3	bc868b1e92644df1483f039ff7c63492274347af	19,3766670227
4	2e4cb7596af3705fa19339c15ffc6371e2892d40	46,6699981689
5	ff15ecdef3d3d658c860ba3ecc1ea8a81bcb262	55,2962493896
6	51173df060d51acabc911ebc5425097534caf15d	114,9416707357
7	851f44cc099987f008b0a1ecb95ffbaf1de92c1	32,9866663615
8	2964e4f55fba91d3f3bfa2cb1e3a3ade3c3b86cb	76,4325027466
9	aa08c1925d8219ed8e12ef3a3952ef4d112caab8	34,9700012207
10	8c3cbadfead76331fe010eca8281f73b226cd99f	29,2362499237
11	e01ac021ca24a2bd32d2fc40fe2b8eb872abf3f5	16,6275005341
12	1f006750cb6cb8129a17fd20ec12d0f62361f614	79,371140951
13	2764828ed6e9e9af7fe8189e99b37bd034f38459	49,5449981689

6. For each booking, what is the total transaction value?

Assumption:

Insurance price = additional_travel_insurance + additional_deductible_insurance

-- 6. For each booking, what is the total transaction value?

Grid	ABC booking_id	123 total_transaction_value
1	3498b368819e9dd3dbcb2774e3228d9a6df0d733	79,9499969482
2	08efe051fbceb577fe579b2cf005eed8d53fe0e0	311,3500061035
3	bc868b1e92644df1483f039ff7c63492274347af	80,6300010681
4	2e4cb7596af3705fa19339c15ffc6371e2892d40	560,0399780273
5	ff15ecdef3d3d658c860ba3ecc1ea8a81bcb262	442,3699951172
6	51173df060d51acabc911ebc5425097534caf15d	689,6500244141
7	851f44cc099987f008b0a1ecb95ffbaf1de92c1	98,9599990845
8	2964e4f55fba91d3f3bfa2cb1e3a3ade3c3b86cb	305,7300109863
9	aa08c1925d8219ed8e12ef3a3952ef4d112caab8	104,9100036621
10	8c3cbadfead76331fe010eca8281f73b226cd99f	233,8899993896
11	e01ac021ca24a2bd32d2fc40fe2b8eb872abf3f5	66,5100021362
12	1f006750cb6cb8129a17fd20ec12d0f62361f614	1,428,6800537109

7. For each booking, what is the net revenue?

Assumption:

Insurance Commission = additional_travel_insurance_comission +
additional_deductible_insurance_comission

The screenshot shows a MySQL Workbench environment. On the left, there's a sidebar with icons for schema, table, view, and procedure. The main area has a SQL editor window with the following content:

```
-- 7. For each booking, what is the net revenue?  
...  
⑤ SELECT booking_id,  
       (booking_supplier_commission + additional_travel_insurance_commission + additional_deductible_insurance_commission)  
     AS total_net_revenue  
  FROM `aws-silvertours-link`.probetag.bookings_trial_day;
```

Below the SQL editor is a results grid titled "Results 1". The grid has two columns: "booking_id" and "total_net_revenue". The data is as follows:

Record	booking_id	total_net_revenue
1	3498b368819e9dd3dbcb2774e3228d9a6df0d733	14,8499996662
2	08efe051fbceb577fe579b2cf005eed8d53fe0e0	60,3799991608
3	bc868b1e92644df1483f039ff7c63492274347af	17,8500003815
4	2e4cb7596af3705fa19339c15ffc6371e2892d40	109,2099990845
5	ff15ecdef3d3d658c860ba3ecc1ea8a81bcb262	79,6299972534
6	51173df060d51acabc911ebc5425097534caf15d	117,2399978638
7	851f44cc099987f008b0a1ecb95ffbaf1de92c1	16,8199996948
8	2964e4f55fba91d3f3bfa2cb1e3a3ade3c3b86cb	55,0299987793
9	aa08c1925d8219ed8e12ef3a3952ef4d112caab8	20,9799995422
10	8c3cbadfead76331fe010eca8281f73b226cd99f	45,6100006104
11	e01ac021ca24a2bd32d2fc40fe2b8eb872abf3f5	11,9899997711
12	1f006750cb6cb8129a17fd20ec12d0f62361f614	257,1600036621
13	2764828ed6e9e9af7fe8189e99b37bd034f38459	17,8400001526

1b- Email marketing campaigns

Following query creates a new table in the database, the table is shown below.

```
-- Creating new Table
CREATE TABLE IF NOT EXISTS probetag.campaigns_trial_day (
    Campaign_ID string,
    Date DATE,
    Number_of_Bookings INT,
    Number_of_Emails_Send INT,
    Number_of_Emails_Open INT,
    Number_of_Emails_Click INT,
    Number_of_Sessions INT,
    Number_of_Transactions INT,
    Number_of_Bookings_with_Transactions INT,
    Total_Transaction_Values Float64,
    Net_Revenue Float64
);

-- Inserting values into the new table
INSERT INTO probetag.campaigns_trial_day (
    Campaign_ID,
    Date,
    Number_of_Bookings,
    Number_of_Emails_Send,
    Number_of_Emails_Open,
    Number_of_Emails_Click,
    Number_of_Sessions,
    Number_of_Transactions,
    Number_of_Bookings_with_Transactions,
    Total_Transaction_Values,
    Net_Revenue
)
SELECT
    e.email_id AS Campaign_ID,
    CAST(MAX(e.email_created) AS DATE) AS Date,
    COUNT(DISTINCT b.booking_id) AS Number_of_Bookings,
    COUNT(e.customer_id) AS Number_of_Emails_Send,
    COUNT(DISTINCT e.email_open) AS Number_of_Emails_Open,
    COUNT(DISTINCT e.email_click) AS Number_of_Emails_Click,
    SUM(ga.ga_sessions) AS Number_of_Sessions,
    SUM(ga.ga_transactions) AS Number_of_Transactions,
    SUM(CASE WHEN ga.ga_transactions > 0 THEN ga.ga_transactions END) AS Number_of_Bookings_with_Transactions,
    SUM(ga.ga_transactions_revenue) AS Total_Transaction_Values,
    SUM(CASE WHEN ga.ga_transactions > 0 THEN ga.ga_transactions_revenue END) AS Net_Revenue
FROM probetag.emails_trial_day AS e
INNER JOIN probetag.bookings_trial_day b
    ON b.customer_id = e.customer_id
    AND b.booking_is_cancelled = FALSE
INNER JOIN probetag.google_analytics_trial_day AS ga
    ON e.email_id = ga.ga_campaigns
WHERE (e.email_open IS NOT NULL OR e.email_click IS NOT NULL)
    AND (DATE_DIFF(DATE(e.email_open), DATE(b.booking_creation), DAY) + 1 <= 7
    OR DATE_DIFF(DATE(e.email_click), DATE(b.booking_creation), DAY) + 1 <= 7)
GROUP BY e.email_id;
```

A	B	C	D	E	F	G	H	I	J	K
Campaign_ID	Date	Number_of_Bookings	Number_of_Emails_Send	Number_of_Emails_Open	Number_of_Emails_Click	Number_of_Sessions	Number_of_Transactions	Number_of_Bookings_with_Transactions	Total_Transaction_Values	Net_Revenue
38948	10/28/22	6	138	6	0	192	12	12	1196.219971	1196.219971
27669	11/18/22	2	238	2	0	298	26	26	1479.100012	1479.100012
24467	10/6/22	7	266	7	1	287	14	14	1278.269964	1278.269964
19994	11/18/22	1	112	1	1	137	14	14	1568.279994	1568.279994
39899	11/11/22	161	62088	191	38	272033	18507	18507	1106772.318	1106772.318

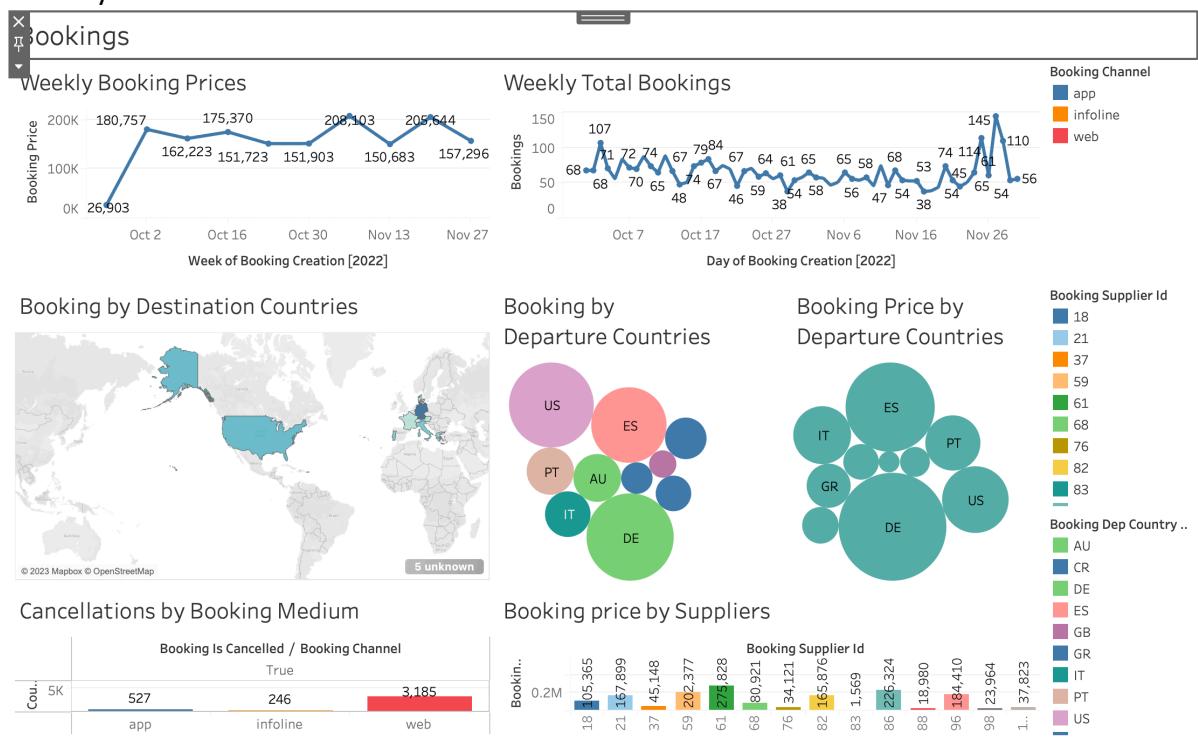
Task 2: Reporting

Dashboard 1

I have very little experience with Tableau but I have very good experience with Google Looker studio, Power BI and Metabase. I have attached some of my Power BI reporting samples along with this file. I have created a basic dashboard for the Bookings data. I am quite sure that I can learn Tableau very quickly as I just need to learn its basics and features.

Link to dashboard:

https://public.tableau.com/app/profile/asad.tariq/viz/Trial_16871159297540/Dashboard1?publish=yes



Task 3. Helping your colleague

Here is the fixed query. The changes are as follows,

- Removed nested Cases.
- Corrected the syntax for the COALESCE function to ensure proper handling of NULL values.

```
SELECT
  COALESCE(
    CASE
      WHEN s.dataset = 'analytics_abc' AND s.event_date::DATE >= '2023-03-27' THEN
        CASE
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) ILIKE '%es%' THEN 'floyt.com/es'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) ILIKE '%it%' THEN 'floyt.com/it'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) ILIKE '%nl%' THEN 'floyt.com/nl'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) LIKE 'es' THEN 'floyt.com/es' -- to cover http://floyt.com/es
          -- to cover http://floyt.com/es-es or floyt.com/es-it etc
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2), '-') LIKE 'es' THEN 'floyt.com/es'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) LIKE 'it' THEN 'floyt.com/it' -- to cover floyt.com/it
          -- to cover http://floyt.com/it-it, it-es, etc
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2), '-') LIKE 'it' THEN 'floyt.com/it'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) LIKE 'nl' THEN 'floyt.com/nl' -- to cover floyt.com/nl
          -- to cover floyt.com/nl-nl, nl-es, etc
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2), '-') LIKE 'nl' THEN 'floyt.com/nl'
          ELSE NULL
        END
      ELSE 'floyt.com/es'
    END,
    am.app_market,
    NULL
  ) AS market,
  COALESCE(
    CASE
      WHEN s.dataset = 'analytics_abc' AND s.event_date::DATE >= '2023-03-27' THEN
        CASE
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) ILIKE '%es%' THEN 'es'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) ILIKE '%it%' THEN 'it'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) ILIKE '%nl%' THEN 'nl'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) LIKE 'es' THEN 'es' -- to cover http://floyt.com/es
          -- to cover http://floyt.com/es-es or floyt.com/es-it etc
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2), '-') LIKE 'es' THEN 'es'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) LIKE 'it' THEN 'it' -- to cover floyt.com/it
          -- to cover http://floyt.com/it-it, it-es, etc
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2), '-') LIKE 'it' THEN 'it'
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2)) LIKE 'nl' THEN 'nl' -- to cover floyt.com/nl
          -- to cover floyt.com/nl-nl, nl-es, etc
          WHEN lower(split_part(split_part(landing_page, '/', 2), '/', 2), '-') LIKE 'nl' THEN 'nl'
          ELSE NULL
        END
      ELSE 'es'
    END,
    ap.app_language,
    NULL
  ) AS "language"
```