Cruisebound

Assignment - Data Analyst

Thank you for your interest in the Data Analyst position. As the next step of our process we would like to invite you to complete the following take-home assignment.

Part A: SQL

Using the table schemas provided in the appendix, provide SQL queries of any variant to:

1. List all upcoming sailings that are longer than 7 days in duration
2. List the top 10 sailings that call at the highest number of distinct ports
3. List the total number of reservations and percentage of reservations made that are refundable each month
4. List the percentage of sailings that are open-jaw vs. closed-loop. A sailing i s classified as *open-jaw* if it begins and ends at different ports, and *closed-loop* if it begins and ends at the same port.

Part B: Exploratory Analysis

Here we work with an external dataset from a Portuguese bank's direct marketing campaign. Download bank.zip from the UC Irvine ML Repository, refer to `bank-names.txt` for information about the variables and data collection, and use the `bank-full.csv` dataset.

Phone calls are labor intensive. Suppose you were asked to provide actionable insights to help the marketing team optimize the use of its limited resources to target clients who are most likely to subscribe to a term deposit (`y='yes'`). With this business objective in mind, perform an exploratory analysis of the dataset and answer the following questions:

1. What actionable insights would you share with the marketing team?
2. What additional data would you like to collect to verify and validate your recommendations?

Deliverables

Please submit your work by emailing us a link to a private GitHub repository containing:

- A .sql file containing your answers to Part A
- Any files used to answer Part B (e.g. .sql, .py, .r)
- A .pdf report containing explanations & recommendations for your implementation of Part B
- Any additional files (e.g. Python/R, Jupyter notebooks) showing your work for Part A or B

Appendix: Data Schema

Table: **partner_marketing_report_YYYY_MM_DD**
(each row represents the performance of one ad for a unique cruise ship on a unique date)

Column	**Data Type**	**Example**	**Comment**	**Nullable**
Date	DATE	2023-12-05	Date of billing	False
CruiseLine	VARCHAR	Celebrity Cruises	Name of cruise line company	False
Ship	VARCHAR	Celebrity Apex	Name of the physical cruise ship	False
MobileClicks	INT	38	Number of clicks from a mobile device	False
DesktopClicks	INT	26	Number of clicks from a desktop device	False
MobileCost	FLOAT	6.16	Cost of mobile advertising measured in BillingCurrency	True
DesktopCost	FLOAT	24.53	Cost of desktop advertising measured in BillingCurrency	True
BillingCurrency	VARCHAR	USD	Currency we are billed in	False
BillingTimeZone	VARCHAR	Eastern Standard Time	Timezone corresponding to date and amount we were billed	False

Table: **sailing**
(each row represents a “sailing” or cruise trip that a customer can book)

Column	Data Type	Example	Comment	Nullable
id	INT	20491	Unique ID of a sailing	False
name	VARCHAR	Mexican Riviera	Name of the sailing	False
description	VARCHAR	On this convenient San Diego round trip Mexican Riviera	Long-form text advertising the sailing	False
line_name	VARCHAR	Holland America	Name of the cruise line company operating this sailing	False
ship_name	VARCHAR	Koningsdam	Name of the physical cruise ship	False
start_date	DATE	2024-03-06	Date at which the cruise sets sail	False
end_date	DATE	2024-03-13	Date at which the cruise completes its trip	False

Table: **itinerary_destination**
(each row represents a destination on the cruise’s itinerary, which can be at sea or at a port)

Column	Data Type	Example	Comment	Nullable
id	INT	98997	Unique ID of an itinerary destination	False
sailing_id	INT	20491	Can be joined to to sailing.id	False
order_id	INT	4	Describes the destination’s order in the sailing’s itinerary	False
arrive_date	DATE	2022-03-09	Date at which the cruise arrives at this destination	False
depart_date	DATE	2022-03-09	Date at which the cruise departs this destination	False
name	VARCHAR	Puerto Vallarta	Name of the destination	False
description	VARCHAR	Unique compared to Acapulco, Cancún…	Long-form text advertising the destination	True

Column	Data Type	Example	Comment	Nullable
type	VARCHAR	port	Can be either “port” if calling at a port, or “”sailing” if at sea	False

Table: **reservation**
(each row represents a reservation made by a customer)

Column	Data Type	Example	Comment	Nullable
id	INT	999	Unique ID of a reservation	False
user_id	INT	1	ID of the user making the reservation	False
date_created	DATE	2021-09-05	Date on which the reservation was made	False
sailing_id	INT	20491	Maps to sailing.id	False
cabin_code	VARCHAR	SA	Indicated room type	True
cabin_name	VARCHAR	Neptune	Name of the cabin type	True
price	FLOAT	4000	Amount paid by customer to Cruisebound	False
cost	FLOAT	3000	Amount paid by Cruisebound to cruise line provider	True
refundable	BOOL	1	1 if refundable, 0 if non-refundable	False