HOTEL BOOKING ANALYSIS - SQL QUERIES

TravClan Business Analyst Assignment

SECTION 1: DATABASE SETUP & TABLE CREATION

Create database

```
CREATE DATABASE hotel_booking_db;

USE hotel_booking_db;
```

Drop table if exists

```
DROP TABLE IF EXISTS hotel_bookings;
```

Create table structure

```
CREATE TABLE hotel_bookings (
customer_id INT,
    property_id INT,
    city VARCHAR(100),
    star_rating INT,
    booking_date DATE,
    check_in_date DATE,
    check_out_date DATE,
    room_type VARCHAR(50),
    num_rooms_booked INT,
    stay_type VARCHAR(50),
    booking_channel VARCHAR(50),
    booking_value DECIMAL(10,2),
    costprice DECIMAL(10,2),
    markup DECIMAL(10,2),
    selling_price DECIMAL(10,2),
    payment_method VARCHAR(50),
    refund_status VARCHAR(10),
    refund_amount DECIMAL(10,2),
    channel_of_booking VARCHAR(50),
    booking_status VARCHAR(50),
    travel_date DATE,
    cashback DECIMAL(10,2),
    coupon_redeem DECIMAL(10,2),
    coupon_used VARCHAR(10),
    length_of_stay INT,
    is_weekend_checkin BOOLEAN,
```

Additional calculated columns

```
    actual_length_stay INT,
    profit_margin DECIMAL(10,2),
    profit_margin_pct DECIMAL(10,4),
    is_cancelled INT,
    is_refunded INT,
    has_coupon INT,
    net_revenue DECIMAL(10,2)
);
```

Load data using appropriate method for your database

Examples - LOAD DATA INFILE 'Hotel_Bookings_Clean.csv' INTO TABLE hotel_bookings;

SECTION 2: DATA QUALITY CHECKS

Check total records

```
SELECT
    'Total Records' AS metric,
    COUNT(*) AS value
FROM hotel_bookings;
```

Check for NULL values in key columns

```
SELECT
    'Null Check' AS analysis,
    SUM(CASE WHEN booking_date IS NULL THEN 1 ELSE 0 END) AS null_booking_date,
    SUM(CASE WHEN check_in_date IS NULL THEN 1 ELSE 0 END) AS null_check_in,
    SUM(CASE WHEN check_out_date IS NULL THEN 1 ELSE 0 END) AS null_check_out,
    SUM(CASE WHEN selling_price IS NULL THEN 1 ELSE 0 END) AS null_selling_price
FROM hotel_bookings;
```

Check for duplicates

```
SELECT
    customer_id,
    property_id,
    booking_date,
    COUNT(*) AS duplicate_count
FROM hotel_bookings
GROUP BY customer_id, property_id, booking_date
HAVING COUNT(*) > 1;
```

Check date consistency

```
SELECT
    COUNT(*) AS inconsistent_dates
FROM hotel_bookings
WHERE check_in_date > check_out_date OR booking_date > check_in_date;
```

SECTION 3: KEY BUSINESS METRICS

Overall Business Metrics

```
SELECT
    COUNT(*) AS total_bookings,
    COUNT(DISTINCT customer_id) AS unique_customers,
    COUNT(DISTINCT property_id) AS unique_properties,
    SUM(selling_price) AS total_revenue,
    SUM(costprice) AS total_cost,
    SUM(selling_price - costprice) AS total_profit,
    AVG(selling_price - costprice) AS avg_profit_per_booking,
    AVG(CASE WHEN costprice > 0 THEN ((selling_price - costprice) / costprice) * 100 END) AS avg_profit_margin_pct,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS total_cancellations,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
    SUM(refund_amount) AS total_refunds,
    AVG(DATEDIFF(check_out_date, check_in_date)) AS avg_stay_length,
    SUM(cashback) AS total_cashback_paid,
    SUM(coupon_redeem) AS total_coupon_value
FROM hotel_bookings;
```

SECTION 4: CANCELLATION ANALYSIS

Cancellation Rate by Booking Channel

```
SELECT
    booking_channel,
    COUNT(*) AS total_bookings,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS cancelled_bookings,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
    SUM(selling_price) AS total_revenue,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN selling_price ELSE 0 END) AS lost_revenue
FROM hotel_bookings
GROUP BY booking_channel
ORDER BY cancellation_rate_pct DESC;
```

Cancellation Rate by Room Type

```
SELECT
    room_type,
    COUNT(*) AS total_bookings,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS cancelled_bookings,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
    AVG(selling_price) AS avg_booking_value
FROM hotel_bookings
GROUP BY room_type
ORDER BY cancellation_rate_pct DESC;
```

Cancellation Rate by Star Rating

```
SELECT
    star_rating,
    COUNT(*) AS total_bookings,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS cancelled_bookings,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
    SUM(selling_price) AS total_revenue
FROM hotel_bookings
GROUP BY star_rating
ORDER BY star_rating DESC;
```

Cancellation Rate: Weekend vs Weekday

```
SELECT
    CASE WHEN is_weekend_checkin = 1 THEN 'Weekend' ELSE 'Weekday' END AS checkin_type,
    COUNT(*) AS total_bookings,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS cancelled_bookings,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct
FROM hotel_bookings
```

```
GROUP BY is_weekend_checkin;
```

Cancellation by Lead Time

```
SELECT
    CASE
        WHEN DATEDIFF(check_in_date, booking_date) BETWEEN 0 AND 7 THEN '0-7 days'
        WHEN DATEDIFF(check_in_date, booking_date) BETWEEN 8 AND 30 THEN '8-30 days'
        WHEN DATEDIFF(check_in_date, booking_date) BETWEEN 31 AND 90 THEN '31-90 days'
        ELSE '90+ days'
    END AS lead_time_bucket,
    COUNT(*) AS total_bookings,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
    AVG(selling_price) AS avg_booking_value
FROM hotel_bookings
WHERE check_in_date IS NOT NULL AND booking_date IS NOT NULL
GROUP BY lead_time_bucket
ORDER BY cancellation_rate_pct DESC;
```

SECTION 5: REVENUE ANALYSIS

Revenue by Booking Channel

```
SELECT
    booking_channel,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_revenue_per_booking,
    SUM(selling_price - costprice) AS total_profit,
    AVG(selling_price - costprice) AS avg_profit_per_booking,
    ROUND(AVG(CASE WHEN costprice > 0 THEN ((selling_price - costprice) / costprice) * 100 END), 2) AS avg_profit_margin_pct
```

```
FROM hotel_bookings
WHERE booking_status != 'Cancelled'
GROUP BY booking_channel
ORDER BY total_revenue DESC;
```

Top 10 Cities by Revenue

```
SELECT
    city,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_revenue_per_booking,
    SUM(selling_price - costprice) AS total_profit,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct
FROM hotel_bookings
GROUP BY city
ORDER BY total_revenue DESC
LIMIT 10;
```

Revenue by Star Rating

```
SELECT
    star_rating,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_revenue_per_booking,
    SUM(selling_price - costprice) AS total_profit,
    ROUND(AVG((selling_price - costprice) / selling_price) * 100, 2) AS avg_margin_pct
FROM hotel_bookings
WHERE booking_status != 'Cancelled'
GROUP BY star_rating
ORDER BY star_rating DESC;
```

Revenue by Stay Type (Leisure vs Business)

```
SELECT
    stay_type,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_revenue_per_booking,
    AVG(DATEDIFF(check_out_date, check_in_date)) AS avg_stay_length,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct
FROM hotel_bookings
GROUP BY stay_type;
```

SECTION 6: TEMPORAL ANALYSIS

Monthly Booking Trends

```
SELECT
    DATE_FORMAT(booking_date, '%Y-%m') AS booking_month,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_booking_value,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
    SUM(selling_price - costprice) AS total_profit
FROM hotel_bookings
GROUP BY booking_month
ORDER BY booking_month;
```

Day of Week Analysis

```
SELECT
    DAYNAME(booking_date) AS day_of_week,
    COUNT(*) AS total_bookings,
    AVG(selling_price) AS avg_booking_value,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS
cancellation_rate_pct
FROM hotel_bookings
GROUP BY DAYNAME(booking_date), DAYOFWEEK(booking_date)
ORDER BY DAYOFWEEK(booking_date);
```

Quarterly Performance

```
SELECT
    YEAR(booking_date) AS booking_year,
    QUARTER(booking_date) AS booking_quarter,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    SUM(selling_price - costprice) AS total_profit,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS
cancellation_rate_pct
FROM hotel_bookings
GROUP BY booking_year, booking_quarter
ORDER BY booking_year, booking_quarter;
```

SECTION 7: CUSTOMER SEGMENTATION

High Value Customers (Top 10%)

```
WITH customer_value AS (
    SELECT
        customer_id,
        COUNT(*) AS total_bookings,
        SUM(selling_price) AS total_spent,
        AVG(selling_price) AS avg_booking_value,
        ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS
cancellation_rate
    FROM hotel_bookings
    GROUP BY customer_id
),
percentiles AS (
    SELECT PERCENTILE_CONT(0.90) WITHIN GROUP (ORDER BY total_spent) AS top_10_pct_threshold
    FROM customer_value
)
SELECT
    cv.customer_id,
    cv.total_bookings,
    cv.total_spent,
    cv.avg_booking_value,
    cv.cancellation_rate,
    'High Value' AS customer_segment
FROM customer_value cv, percentiles p
WHERE cv.total_spent >= p.top_10_pct_threshold
ORDER BY cv.total_spent DESC;
```

Customer Retention Analysis

```
SELECT
    customer_id,
    COUNT(*) AS total_bookings,
    MIN(booking_date) AS first_booking,
    MAX(booking_date) AS last_booking,
    DATEDIFF(MAX(booking_date), MIN(booking_date)) AS customer_lifetime_days,
    SUM(selling_price) AS lifetime_value
FROM hotel_bookings
GROUP BY customer_id
HAVING COUNT(*) > 1
ORDER BY total_bookings DESC
LIMIT 100;
```

SECTION 8: PAYMENT METHOD ANALYSIS

Payment Method Performance

```
SELECT
   payment_method,
   COUNT(*) AS total_bookings,
   SUM(selling_price) AS total_revenue,
   AVG(selling_price) AS avg_transaction_value,
   ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
   ROUND(AVG(CASE WHEN refund_status = 'Yes' THEN 1 ELSE 0 END) * 100, 2) AS refund_rate_pct
FROM hotel_bookings
GROUP BY payment_method
ORDER BY total_bookings DESC;
```

SECTION 9: ROOM TYPE PROFITABILITY

Room Type Profitability Analysis

```
SELECT
   room_type,
   COUNT(*) AS total_bookings,
   SUM(selling_price) AS total_revenue,
   AVG(selling_price) AS avg_revenue,
   SUM(selling_price - costprice) AS total_profit,
   AVG(selling_price - costprice) AS avg_profit,
   ROUND(AVG((selling_price - costprice) / selling_price) * 100, 2) AS profit_margin_pct,
   ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct,
   AVG(DATEDIFF(check_out_date, check_in_date)) AS avg_stay_length
FROM hotel_bookings
GROUP BY room_type
```

```
ORDER BY total_profit DESC;
```

SECTION 10: PROPERTY PERFORMANCE

Top 20 Properties by Revenue

```
SELECT
    property_id,
    city,
    star_rating,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_revenue_per_booking,
    SUM(selling_price - costprice) AS total_profit,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS
cancellation_rate_pct
FROM hotel_bookings
GROUP BY property_id, city, star_rating
ORDER BY total_revenue DESC
LIMIT 20;
```

SECTION 11: PROMOTIONAL EFFECTIVENESS

Coupon Usage Analysis

```
SELECT
    CASE WHEN coupon_used = 'Yes' THEN 'With Coupon' ELSE 'Without Coupon' END AS coupon_status,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    AVG(selling_price) AS avg_booking_value,
    SUM(coupon_redeem) AS total_discount_given,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS
cancellation_rate_pct
```

```
FROM hotel_bookings
GROUP BY coupon_status;
```

Cashback Impact

```
SELECT
    CASE
        WHEN cashback = 0 THEN 'No Cashback'
        WHEN cashback > 0 AND cashback <= 5 THEN 'Low (0-5)'
        WHEN cashback > 5 AND cashback <= 15 THEN 'Medium (5-15)'
        ELSE 'High (15+)'
    END AS cashback_tier,
    COUNT(*) AS total_bookings,
    AVG(selling_price) AS avg_booking_value,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct
FROM hotel_bookings
GROUP BY cashback_tier
ORDER BY cashback_tier;
```

SECTION 12: ADVANCED INSIGHTS

Channel Performance with Profitability

```
SELECT
    booking_channel,
    COUNT(*) AS bookings,
    SUM(selling_price) AS revenue,
    SUM(selling_price - costprice) AS profit,
    ROUND(AVG((selling_price - costprice) / selling_price) * 100, 2) AS margin_pct,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancel_rate,
    SUM(coupon_redeem) AS total_coupons,
    SUM(cashback) AS total_cashback,
    SUM(selling_price - costprice - coupon_redeem - cashback) AS net_profit
FROM hotel_bookings
GROUP BY booking_channel
ORDER BY net_profit DESC;
```

Repeat Customer Analysis

```
SELECT
    customer_id,
    COUNT(*) AS booking_count,
    SUM(selling_price) AS total_value,
    AVG(selling_price) AS avg_booking_value,
    STRING_AGG(DISTINCT city, ', ') AS cities_visited,
    STRING_AGG(DISTINCT booking_channel, ', ') AS channels_used
FROM hotel_bookings
GROUP BY customer_id
HAVING COUNT(*) >= 3
ORDER BY booking_count DESC, total_value DESC
LIMIT 50;
```

City-wise Star Rating Distribution

```
SELECT
    city,
    star_rating,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    ROUND(AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) * 100, 2) AS cancellation_rate_pct
FROM hotel_bookings
GROUP BY city, star_rating
ORDER BY city, star_rating DESC;
```

SECTION 13: EXPORT VIEWS FOR POWER BI

Create view for Power BI Dashboard

```
CREATE OR REPLACE VIEW vw_powerbi_main AS
SELECT
    customer_id,
    property_id,
    city,
    star_rating,
    booking_date,
    check_in_date,
    check_out_date,
    room_type,
    stay_type,
    booking_channel,
    booking_value,
    selling_price,
    costprice,
    selling_price - costprice AS profit,
    CASE WHEN costprice > 0 THEN ((selling_price - costprice) / costprice) * 100 ELSE 0 END AS profit_margin_pct,
    payment_method,
    booking_status,
    CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END AS is_cancelled,
    refund_status,
    refund_amount,
    cashback,
    coupon_redeem,
    coupon_used,
    DATEDIFF(check_out_date, check_in_date) AS length_of_stay,
    DATEDIFF(check_in_date, booking_date) AS lead_time,
    is_weekend_checkin,
    MONTH(booking_date) AS booking_month,
    YEAR(booking_date) AS booking_year,
    QUARTER(booking_date) AS booking_quarter,
    DAYNAME(booking_date) AS booking_day_of_week
FROM hotel_bookings;
```

Create aggregated metrics view

```
CREATE OR REPLACE VIEW vw_powerbi_metrics AS
SELECT
    DATE_FORMAT(booking_date, '%Y-%m') AS month_year,
    booking_channel,
    city,
    star_rating,
    room_type,
    COUNT(*) AS total_bookings,
    SUM(selling_price) AS total_revenue,
    SUM(selling_price - costprice) AS total_profit,
    AVG(selling_price) AS avg_booking_value,
    SUM(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS cancelled_bookings,
    AVG(CASE WHEN booking_status = 'Cancelled' THEN 1 ELSE 0 END) AS cancellation_rate
FROM hotel_bookings
GROUP BY month_year, booking_channel, city, star_rating, room_type;
```