Greenlight Planet sells lamps and TVs that are repaid over time by its customers. It collects data in real time to track whether customers' payments are made on time and how much of the amount owed has been paid so far.

If a customer falls behind with their payment, then their lamp or TV stops functioning until a new payment is made. This is reflected in the database by labeling the customer's account status as 'disabled'.
It is common to calculate the ratio of the number of lamps that are disabled over the total number of lamps of any significant group of customers (by country, area, date or month when the customers were registered, etc.). This ratio is known as '*disabled_rate*' and is an important monitoring metric for the health of the company.

When purchasing a lamp, a customer pays a deposit and is thereafter expected to make a certain agreed upon daily payment until the price of the unit is paid in full after the repayment period has passed. This allow us to define the following metrics

- Upfront_price: initial deposit made when purchasing the lamp
- Price_per_day: amount to be paid for one day of utilization of the product
- Unlock_price: total price to be paid to own the product i.e. price of the product
- Follow_on_price: unlock_price - upfront_price
- Follow_on_paid: sum of the amounts paid by the customer
- Follow_on_revenue: sum of the follow_on_paid for multiple customers
- Follow_on_revenue_realization (FRR): follow_on_paid / follow_on_price or follow_on_revenue / sum(follow_on_price) for multiple customers. The FRR is monitored on a daily basis to ensure the health of the portfolio of customers of the company. Anything short of 100% represents a loss that needs to be offset one way or another.

The price_per_day can be paid daily, or the customer can make larger payments that allow the lamp to remain enabled for a number of days equivalent to the amount paid (amount / price_per_day). Customers are expected to have made a payment of price_per_day for each day after the lamp/TV is registered, but often times a payment is missed and the lamp/TV becomes disabled for one or a few days. Note the following relationships between the metrics previously defined:
- follow_on_paid = sum(amounts)
- days_of_light_paid = follow_on_paid / price_per_day
- number_of_days_disabled = number_of_days_since_registration - days_of_light_paid
- Be careful not to calculate follow_on_paid = number_of_days_since_registration * price_per_day, because number_of_days_since_registration doesn't necessarily equates to the number of times the price_per_day was paid

Greenlight Planet regularly runs pilots to test new prices, products, or systems for its operations. When doing so, one or a few areas of operations are selected to try a new methodology, called '*treatment*'. We then want to compare the treatment's key performance indicators (KPIs) against

those of the rest of the country the pilot areas pertain to, as well as against the performance of a benchmark country considered to be healthy. The main KPIs followed are the FRR and the disabled_rate described above.

A pilot was recently run and your manager asks you to analyse the pilot data and prepare a powerpoint presentation to outline your findings to the head of the Africa sales department.

S/he provides you with data for customers registered from a few months before the beginning of the pilot until recently. Every 2 weeks the FRR and whether the customer's lamp is disabled or not are recorded and shown in the data provided. Data is provided for Country 1 and 2, for the accounts having received treatment or not. Only some customers from Country 1 have received the treatment, Country 2 is only used for benchmark purposes.
To limit the size of the data, your manager sampled customers from Country 2 included in the dataset used for the analysis. The sample size was chosen in order to have a 0.5% confidence interval with 99% confidence level, see below details about the samples:

Country	is_treatment	Total_count	Sample_count
Country 1	Yes	1350	1350
Country 2	No	138195	45086
Country 1	No	3838	3838

We can generally assume that existing customers are representative of all potential customers, although this might not be entirely true depending on new pricing schemes, new products, and potential new areas of operations.

Your manager asks you to explore each of the following points, and present them in whichever order you see fit:
- Build and display on a graph:
  - The evolution over time of the disabled rates for Country 1 with/without treatment and Country 2
  - The evolution over time of the FRR for Country 1 with/without treatment and Country 2
- Provide a description and analysis of the graphs
- Define and explain how and why hypothesis testing theoretically works to the head of the sales department. Make sure you are both mathematically precise and clear enough as the head of the Africa sales department might not have an extensive statistical background.
- Run a test to prove whether your earlier observations and conclusions might be due to chance or are statistically relevant. Make sure you explain what possible tests could

have been run given the situation and why the one you have chosen is the one that best fits the problem at hand.

- BONUS 1: run a test to prove/find how much of the difference between the metrics average of the treatment/benchmarks cannot be due to chance from a statistical perspective. Make sure to explain the mathematical reasoning.
- BONUS 2: assuming all of the data is stored in one table within the database, write the SQL query that would have allowed you to get the data for the analysis.

Ensure that your presentation is clear for a non-technical audience and make sure to keep them interested throughout.