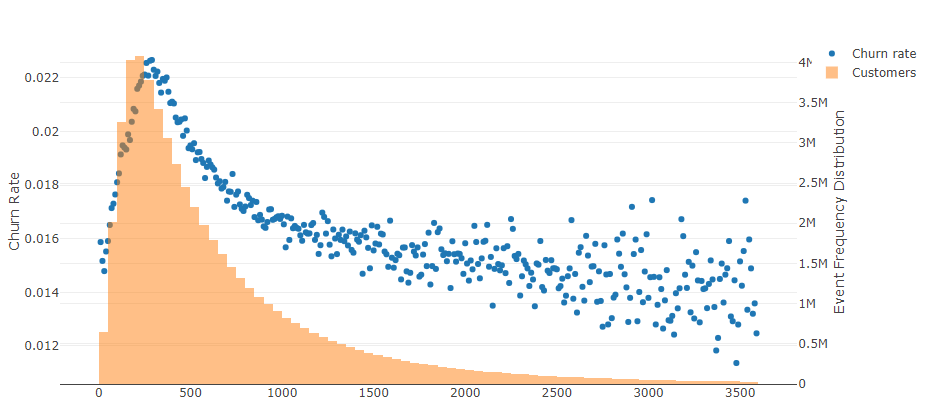
**Part 1: Analysis**

  
Please study the chart above. It provides information on a retail store’s waiting times for their customers. There are several data elements represented:

1. The x-axis is measured in seconds and represents the amount of time existing customers had to wait in the queue of a telecom operator store during their visits. This is rounded to the nearest 10 seconds
2. The blue dots are on the left y-axis and represent the average customer churn rate at the associated waiting time
3. The orange bars are the number of visits in total that experienced a particular waiting time and align to the right y-axis

Now please answer the following questions, taking into account the underlying distribution of events:

1. Please comment on the relationship between waiting times and churn. Would you say that there is a significant relationship that exists?
2. Does this data have any outliers that need to be accounted for? Please elaborate on any decisions.
3. Would you expect waiting time to have a good correlation with churn? Why or why not?
4. Suppose you are tasked with building a statistical model to predict the probability of customer churn with the only input of waiting time.
   1. Which model types are a good candidate for doing this?
   2. What considerations should be made to the waiting time to improve the fit of the model?
   3. How would you test the model fit?

**Part 2: Databases and SQL**

You've been hired to build a social network from the ground up.

"We would like to store a list of users in the database with their corresponding network of friends, groups as well as any messages which are swapped between them. A user consists of a username, password, first name, last name and profile picture. A message is a string. Users can have many friends but usually start off with none or only 1 friend. A user can join or create a group and there can only be 1 creator for any group"

1. Draw up a database schema for the above description with as much detail as required. You may use any tool you like to do this (dbdiagram.io is a free alternative)
2. Demonstrate how you could query your database to power the social network.
   1. A message list of all messages between users X and Y. The list should show the message, the sending user, and their profile pic.
   2. A list of friends of user X.
   3. A list of friends of friends of user X.
   4. A ranked list of the most used words in messages.
3. What is the difference between a LEFT JOIN and an INNER JOIN?
4. What is the difference between a WHERE clause and a HAVING clause?
5. Why do we have indexes?