# Align Technology Take-home challenge questions

## Task 1: Modeling and Insights

Description:

Attached is a dataset for patients’ orthodontic treatment outcomes. Each row represents the treatment outcome of one patient. Below is the description of the variables:

* has\_heard\_ortho: whether a patient is aware of orthodontic treatment before he/she starts the treatment
* patient\_age\_group: patient’s age when treatment is started
  + 0: > 50 years old
  + 1: < 18 years old
  + 2: between 18 and 50
* boosted\_force: the amount of force applied by Align’s proprietary technology to help move teeth
* planned\_treatment: the amount of tooth movements prescribed by Dr to move teeth
* hours\_wear\_aligners: the avg number of hours a patient wears aligners per day
* hours\_wear\_accelerator: the avg number of hours a patient wears an accelerator to help stimulate tooth movements
* counter\_force: the amount of force applied by Align’s proprietary technology to help reduce unwanted tooth movements
* treatment\_outcome: the outcome of a patient’s treatment
  + This is the target variable

Your goals:

1. Create a model to predict “treatment\_outcome” using the data provided. Evaluate the model and try to achieve the best results you can.
2. Explain how the predictor variables impact your target variable.
3. Extract useful insights when exploring the dataset, and think about how those insights might help product development.

Deliverables: Use Python or R to perform prediction or analysis. Choose your way to deliver conclusions and insights, such as using a python/R notebook, PPT or a report.

## Task 2: Sample size

Description:

We recently improved our product design to reduce the aligner breakage rate and want to measure the impact of the improvement.

Question to answer:

How many samples do we need to show a 15% point reduction in breakage rate (before – after =15%) after the product improvement? We can assume 95% confidence level and 90% power.

## Task 3: SQL query

Description:

The database consists of the following tables:

 Create table dbo.customer    (

customer\_id   int          identity primary key clustered

, customer\_name nvarchar(256) not null

)

Create table dbo.purchase\_order     (

purchase\_order\_id    int   identity primary key clustered

, customer\_id          int   not null

, amount              money not null

, order\_date          date  not null

)

Deliverables:

Implement a SQL query for a report that will provide the following information: for each customer output top 5 dates which contain the highest mean amount.

 Possible result:

|  |  |  |
| --- | --- | --- |
| **customer\_name** | **order\_date** | **mean\_amt** |
| Bond, James | 1/10/2011 | 32041 |
| Bond, James | 2/5/2011 | 33229 |
| Bond, James | 3/19/2011 | 30526 |
| Bond, James | 3/25/2011 | 36804 |
| Bond, James | 3/29/2011 | 33545 |
| Dow, Jones | 1/2/2011 | 34674 |
| Dow, Jones | 1/5/2011 | 33128 |
| Dow, Jones | 1/15/2011 | 39672 |
| Dow, Jones | 1/26/2011 | 39939 |
| Dow, Jones | 2/2/2011 | 19912 |
| McCormick, Kenny | 1/22/2011 | 39138 |
| McCormick, Kenny | 2/5/2011 | 31609 |
| McCormick, Kenny | 2/17/2011 | 19982 |
| McCormick, Kenny | 3/19/2011 | 32874 |
| McCormick, Kenny | 3/24/2011 | 34659 |