

## **BBM TEST FOR THE DATA ANALYST POSITION**

For the following questions, please submit your answers and test results in PDF format.

**1.What are the tools that you use on a daily basis for ETL and Data Visualization?**

**2. Write a function (in any language) to reverse word by word in a sentence, a word is any sequence of character separated with space, following are the requirements for the function:**

- It should receive a string.
- It should reverse the string word by word.
- It should consider anything separated by space is a word.
- It should return the reversed sentence.

Examples:

```
reverse("I love data analytics! ");  
=> "analytics! data love I"
```

**3.Write a function (in any language) to measure the minimum number of replacement required to change one string to another, following are the requirements for the function:**

- It should receive two strings.
- It should only measure strings with equal length, otherwise you should return -1.
- It should return the minimum number of replacement to change one string to another.

Examples:

```
measure('GAGCCTACTAACGGGAT', 'CATCGTAATGACGGCCT')  
=> 7  
measure('TAG', 'GAT')  
=> 2  
measure('GGACTGA', 'GGACTGA')  
=> 0  
measure('GGAC', 'GGACTGA')  
=> -1
```

For the following question, please refer to *test\_data.csv*, note that the descriptions of the attributes are in *Attribute Descriptions.xlsx*. The dataset is some hotel room search and booking records from *expedia*. Your task is to import the dataset into a data analytics tools of your choice and answer the following questions. **Note: all questions must be done without modifying the data using excel.**

List the monthly top 5 sites by their total booking from February to April and display their gross during those 3 months using any preferred SQL query language. Please provide the statement and the name of the tool you used.

How are the average star rating and the average review score of branding hotels compare to non-branding hotels? Please also interpret the results.

There are many attributes in the dataset, but not every attribute is useful for determining why customers book the rooms (booking\_bool). Please list your reasoning and select below the attribute(s) you decide to overlook when predicting if a customer is going to book the room:

Date_time	Prop_location_score1	Prop_location_score2	Random_bool
Price_usd	Gorss_booking_usd	Comp#_rate	Comp#_inv
Srch_booking_window	Srch_room_count	Srch_length_of_stay	Position
Visitor_hist_adr_used	Visit_hist_starrating	Prop_review_score	Srch_destination_id