

## Essay Writing

1. Why are you interested in the job?
2. What do you hope to get out of it?
3. What can you uniquely contribute to the team?
4. In the same document, **give a short description of the three accomplishments** (personal or professional) that you are most proud of.
5. Can you describe a project which you led (or were involved) and you had to work with different stakeholders and encountered a difficult situation that threatened delivery of expectations? What was the project, what was the difficult situation, why was it difficult, what did you do that made a difference?
6. Can you describe a routine task that is part of your key result area which you do not like? What is this task, what are the results expected from you, why do you not like it, what do you do to get the task done?
7. Can you describe a situation where you had a task to be completed but did not have all the information you need to deliver a good job? What was the task, what was the situation, how did you go about the challenges?
8. Given the following corporate values: solicitude, excellence, respect, ingenuity with integrity, and teamwork choose one strength that you have and describe a critical incident to show why this was your strength; also, choose one development opportunity for you and describe a critical incident that highlights why this is a development opportunity.

## Coding Test

### About the Task:

Blume is an investment bank and financial services company that has been a prime customer of Import Genius(IG) for the past 5 years. They consume the supply chain descriptive analytics data related to trade across the globe from IG platform which drives some of their critical decisions related to investing, mergers, and acquisitions.

Blume is aiming to create a decision support system with an objective to make their investment decisions better. One of the particular projects for the decision support system is to analyze US companies and their respective products who have been actively importing from various countries across the globe. With that objective, they have made a deal with IG to have all the data related to US imports for the year 2021. They are particularly interested in imports via ocean freight. The data mainly contains the list of all the importers and the relevant attributes associated with it.

You are hired by Blume as a Data Scientist to take the lead on this project. You will be presenting to the investment bankers and stakeholders of the project about the usefulness of the data that was purchased from IG. The objective of the prototype is to:

1. Conduct an exploratory data analysis on the data that you have received
2. Create actionable insights focusing on the descriptive analysis of the provided data.
3. Create a predictive or a machine learning model that can be useful for the decision support system
4. Create a presentation based on your work that'd be presented to the stakeholders

## Instructions:

### Data Description:

There will be three files in the form of CSV that you can access from the google drive folder shared with you. The data description of those files are mentioned below:

1. Main.csv: This is the primary data file which contains the following fields:
  - a. Consignee:
    - i. An object data type where each row is a dict type containing the name and address of the consignee. Consignees in short means importers or the entity who will receive the consignment. They can be Company, institute, person, school etc. In this particular case, we are only interested in companies
    - ii. Data Type: Object
  - b. Calculated TEU:
    - i. TEU stands for Twenty equivalent Unit. It is the measuring unit for calculating the cargo capacity. Read more about it [here](#).
    - ii. Data Type: String
  - c. File Month:
    - i. Month in 2021, represented in the form of a number.
    - ii. Data Type: Int
  - d. Port of Unlading:
    - i. Port of Unloading is a place where a vessel unloads or discharges some or all of its shipments. In this case, the port of unlading should be treated as the ultimate destination of the shipment in the United States.
    - ii. Data Type: Int
  - e. Country of Origin:
    - i. Country from where the shipment is shipped or exported.
    - ii. Data Type: String

- f. HS Codes:
  - i. The Harmonized System(HS) code is a standardized numerical method of classifying traded products.
  - ii. Data Type: String
- 2. Ports.csv: This file contains information on the port of unloading. Can be useful to fetch port names based on the port of unloading number. It contains following fields:
  - a. Port code:
    - i. Port of unloading code in united states
    - ii. Data Type: Int
  - b. Port Name:
    - i. The port name for the respective port of unloading code in United States
    - ii. Data Type: Object
- 3. Company Normalized: This file contains a list of top 10k consignee names mapped with their normalized name. We are particularly interested in the importers mentioned under this file. The importers mentioned in the file are also ranked(only valid for 2021) based on multiple parameters for 2021. It contains following fields:
  - a. Importer Rank:
    - i. Rank of the company in 2021
    - ii. Data Type: Int
  - b. Importer raw name:
    - i. Raw name found in the Main.csv under consignee column for name key.
    - ii. Data Type: Object
  - c. Importer Deduped Name:
    - i. Entity resolved name of the company
    - ii. Data Type: Object

You can fetch the list of HS codes from this [link](#)

### **How to proceed:**

1. You are expected to complete this project using a jupyter notebook
2. You will also be expected to give a presentation on your technical interview day based on your project
3. Upload all you work in the google drive under EXAM folder
4. Your trained models and any other dependencies should also be stored in the same folder for consistencies

**Evaluating parameters for the interview:**

1. Approach towards the problem
2. Insights generated
3. Methodology
4. Statistical Significance
5. Presentation