

## Take Home assignment

### Description

Attached is a CSV file which has the following details about products from various ecommerce sites.

- UUID (Primary Key)
- Price (String)
- Price\_unfiltered (String)
- Product Type
- Category
- Level 1 -> This is just a classification of a product.

CSV FILE -> [DOWNLOAD](#)

**Task:** Use the database to find the following results from this data.

1. **Products without prices**
2. **Count of products** without prices and with prices in **each** Product Type, Category, Level 1
3. **Correct** Product Prices in the **correct format (eg: \$56)** wherever possible and separate them into currency and value columns.
4. List out the categories with average price of product.

**Brownie Points:** Use graphs to depict this data. Use of Cloud Services

### Guidelines:

- Please commit often and with good commit messages. This will allow us to see how you've approached the problem. Don't worry about changing things around often.
- Don't hesitate to ask any questions if you're uncertain about the task or anything else is unclear.
- Some hints: The number of records is large, simply looping over the data is not the solution. Use libraries like Pandas or PySpark to process the data.
- Please submit github link of your assignment after 2 days of sharing the assignment.
- Please make a detailed readme on how to run the code and don't forget to include requirements for PIP.

**Technology to be used: Python (Preferred) or Java**

Send your assignment to [careers@bezen.eco](mailto:careers@bezen.eco) with the title as Data Engineer Assignment. In the email, include your name, phone and LinkedIn/Internshala profile

