

---

# Data Engineer

## Level 2 Test

May 5 2020

---



There are three files shared with you with the following structures:

**FILE NAME:** orders\_test

Column	Description	Relates to
id	Unique transaction ID	
store_id	Unique store ID. May appear more than once in the dataset	store_test.id
customer_id	Unique user ID. May appear more than once in the dataset	customer_test.id
total	Total cost of order	
order_date	Date of Purchase	

**FILE NAME:** store\_test

Feature	Description	Relates to
id	Unique store ID	orders_test.store_id
name	Name of store	
address	Address of store	

**FILE NAME:** customer\_test

Feature	Description	Relates to
id	Unique user ID	orders_test.customer_id
first_name	First Name of the user	
last_name	Last Name of the user	
email	Email Address of the user	

You are required to create two CSV files and email them back to us:

1. An aggregate table showing the total orders and revenue each store had each month. It should have the following columns:

Year (Eg: 2020)

Month (Eg: January)

Store Name

Number of Orders

Total Revenue

2. A list of users who have placed less than 10 orders. It should have the following columns:

First Name

Last Name

Email

Orders Placed by user

### **BONUS POINTS:**

1. Use a MD5 hash to encrypt the emails of the users before converting it to CSV.
2. Upload the CSVs to cloud storage (eg: AWS S3) and share them with us through a publicly accessible URL.

You have until the end of Wednesday, May 6 2020 to make your submission. Good Luck!  
If you require any clarification please reach out to us.