# Take Home Assignment - Senior Data Engineer

## 1 Paddle's Data Engineering Take Home Challenge

### 1.1 Table of Contents:



### 1.2 Take Home Challenge Overview

Thank you for taking time to work on our take home challenge. We appreciate the time and extra effort that it takes to do it. With this take home challenge, we hope to get a glimpse into the way you break down problems and solve them with SQL. Details about the data and the questions are found below in the SQL part of this document. You can find the data in the zip file, which contains multiple csv files. Your solution can be written in any SQL Dialect, please indicate which one you have used.

Please share your solutions in a zip file. Add a file called **sql_answers** with the following files:

- sql_question_1.sql: your SQL code for question 1
- sql_question_2.sql: your SQL code for question 2
- sql_question_3.sql: your SQL code for question 3
- sql_question_4.sql: your SQL code for question 4

We believe that to complete the challenge you'll need around 2 to 3 hours of active development. In these hours, we are not counting time to get set up e.g.reading and understanding the questions.

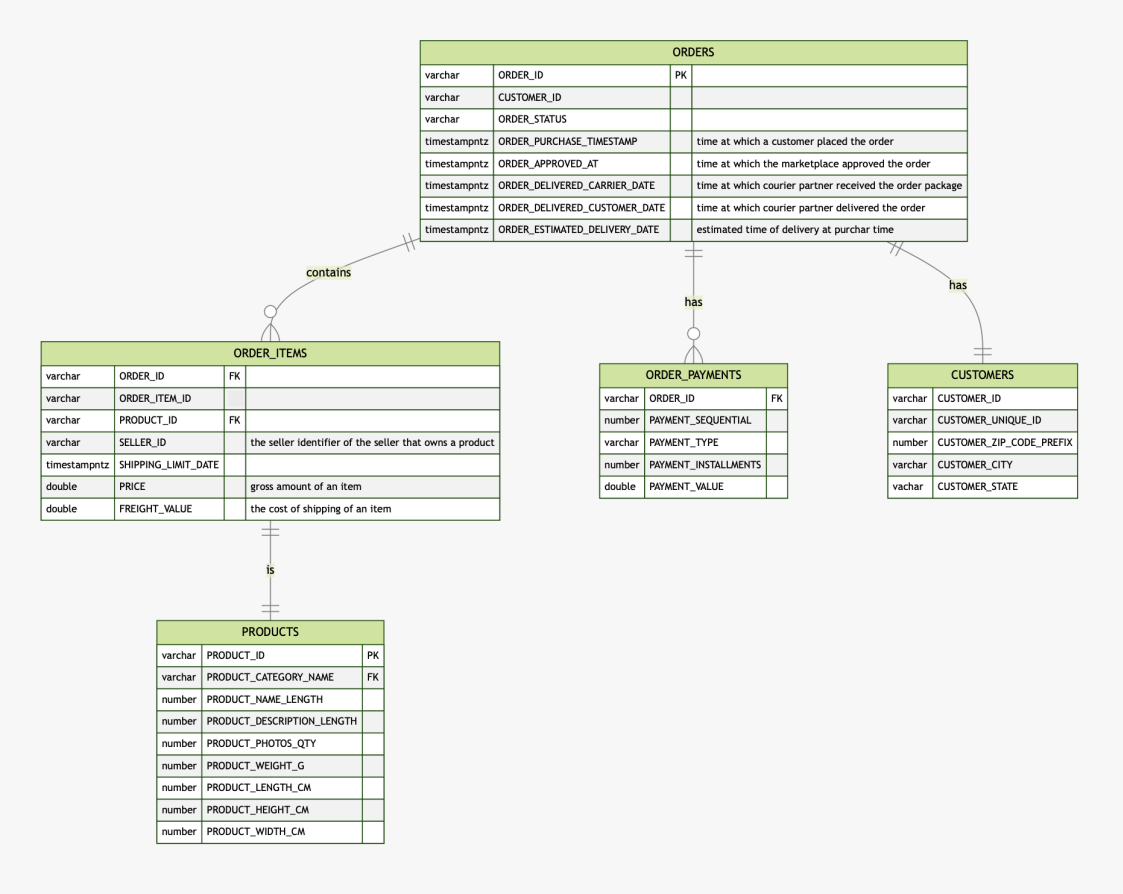
#### 1.2.1 General guidelines

Feel free to add as much commentary as you want or even share a file with your findings. If you get stuck or don't have time to continue coding, feel free to share pseudo-code and a description of the approach you would have taken. Beyond the execution, we want a clear idea of how you understand the problems and how you would potentially optimize or improve your solution.

## 1.3 SQL challenge: ecommerce analytics

### 1.3.1 The data

The SQL challenge is based on e-commerce data: users of an online marketplace can purchase products from multiple sellers. You can see the data model below. The data model is not perfect, if necessary you'll have to derive information from the available tables.




**Database Model / ERD**

### 1.3.2 SQL Question 1

Write some tests that could be used to ensure data accuracy and quality. Propose a few and present them in a SQL file.

### 1.3.3 SQL Question 2

For each month of the year 2017, please calculate the following values:

- monthly active sellers. monthly active seller definition: 25 orders or more in a month avg.
- weekly active sellers. weekly active seller definition: 5 orders or more in a week avg.
- daily active sellers. daily active seller definition: 1 order or more in a day

Present the calculated values for each month in a single row.

### 1.3.4 SQL Question 3

For each of the three product categories that had the most items sold in November 2017, present for the full calendar year of 2017 the following:

- The running sum of their respective total sold gross merchandise value (GMV) using order_items.price on a weekly basis
- Their respective weekly GMV growth rate

### 1.3.5 SQL Question 4

Create a downstream data model which returns all orders that have been shipped and delivered to different cities and states in 2018, which have been paid either with a debit or a credit card, and their value broken down by day. The model should additionally contain the following fields:

- SELLER_ID
- PRODUCT_ID
- ORDER_DELIVERED_CARRIER_DATE