
**Setup**

This exercise is intended to emulate a typical Data Science project at JiffyShirts. You will need to do some data-wrangling, answer a few exploratory questions, build a model and explain your findings over a video interview. You are welcome to use the tools and programming languages you are most comfortable with (e.g. Python/notebooks, R, SQL) and make any assumptions you need. Please limit yourself to no more than 4 hours.

**Background**

JiffyShirts.com is an e-commerce company based in North America with a focus on “soft goods” products like t-shirts, sweaters, and bags. The included datasets emulate the sales of three major products for sale in our store over the last few years. Your goal is to give insight into our sales figures by building a model to forecast total daily revenue.

Details on the datasets:

| **orders.csv** | |
|---|---|
| order_number | (Primary key) A unique identifier for each order placed |
| ordered_at_utc | The UTC time at which an order was placed |
| customer_uuid | A unique identifier for each customer |
| discount | The discount percentage applied to an order (i.e. 0.02 = 2% off the total price of the order) |
| **line_items.csv** | |
| line_item_number | (Primary key) A unique identifier for each line item within an order (e.g. Order containing 5 products – 3 medium red and 2 large blue – would have two line items) |
| order_number | The order associated with each line item |
| units_sold | Number of units on each line item |
| product_id | The product on each line item |
| **products.csv** | |
| product_id | (Primary key) A unique identifier for each product/model version |
| model_number | The general model number of the product |
| description | A description of the model |
| size | Size of the product |
| color | Color of the product |
| selling_price | Selling price of the product (customer pays this) |
| supplier_cost | Supplier price of the product (we pay this) |

**Exercise A**
Import the three csv files into a SQL database and answer the following questions using SQL. You can create the database however you'd like, but please use SQL to solve these questions.

1. How many orders were completed in 2018? (Note: We operate in US/Eastern timezone)
2. How many orders were completed in 2018 containing at least 10 units?
3. How many customers have ever purchased a medium sized sweater with a discount?
4. How profitable was our most profitable month? (Profit = Revenue - Cost)

**Exercise B**
The dataset ends before the year 2019 is over, but we'd like to know how we might perform for the remainder of the year.

1. Build a model to forecast daily revenue for the remainder of 2019. You can use any method you prefer, but please report on your accuracy using a training and test set.
2. Plot both the forecasted and actual revenue on a chart.

**Exercise C**
This question is intended to be open-ended. What other trends do you observe in the dataset? Are there any "nuggets" we could use to enhance our business strategy?

**Exercise D**
Upload your code/answers to github and send a link to sean.malone@jiffyshirts.com