Setup

This exercise is intended to emulate typical data science 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 software 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 are an approximated sample of our sales data. Your goal is to give insight into our sales and come up with some valuable take-aways for the business.

Data Overview

orders.csv	
order_id	(Primary key) A unique identifier for each order placed
customer_uid	(Foreign key) A unique identifier for each customer
order_timestamp	The UTC time at which an order was placed
discount	The discount percentage applied to an order (i.e. 0.02 = 2% off the total price of the order)
ship_state	The US state in where the order was shipped to
shipping_revenue	The customer-paid shipping fee (customer pays this)
shipping_cost	Cost of shipping the order (we pay this)
returned	(T/F) whether the order was returned or not
line_items.csv	
line_item_id	(Primary key) A unique identifier for each line item within an order (e.g. Order containing 5 units – 3 medium red and 2 large blue – would have two line items)
order_id	(Foreign key) The order associated with each line item
quantity	Number of units on each line item
selling_price	Selling price of the product (customer pays this)
supplier_cost	Supplier price of the product (we pay this)
product_category	Product category
color	Color of the product
size	Size of the product
customers.csv	
customer_uid	(Primary key) A unique identifier for each customer
is_business	(T/F) has indicated they are a business
has_account	(T/F) has made user account with JiffyShirts.com
bill_state	Customer's billing address state
acquisition_channel	Marketing channel where the customer was first acquired

Part 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. Here are some helpful guides to loading csv files into a local SQL database:

- Python-based SQLite db - https://stackoverflow.com/questions/41900593/csv-into-sqlite-table-python

OR

- Setting up SQLite db (Mac) - https://razorsql.com/articles/sqlite_mac.html
- Setting up SQLite db (Windows) - https://www.sqlitetutorial.net/download-install-sqlite/
- Importing CSVs into db - https://www.sqlitetutorial.net/sqlite-import-csv/

1. How many orders were completed in 2018? (Note: We operate in US/Eastern time zone)
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?
5. What is the return rate for business vs. non-business customers?

Part B
This question is intended to be open-ended and test your analytical skills alongside your technical skills. We'd like you to use the dataset to build some kind of model that could provide business value. (Don't worry if you don't find anything *actually* valuable, the attempt is what we want to see)

Here are some suggestions:

- **Financial Forecast** -- Forecast our revenue for the next ~30 days after the dataset ends so we can position ourselves with an appropriate promotional strategy.
- **Customer Segmentation** -- How should we segment our customers to align different kinds of marketing messaging and offers?
- **Product Recommendation** -- What products should we recommend to our top customers? Why?
- **Pricing Elasticity** -- Could we change the price of any products (promotionally or permanently) in order to increase sales/profitability?

Feel free to use whatever languages/packages you're most comfortable with, but be prepared to walk-through your code and decisions.

Part C
Upload your code/answers to github and send a link to sean.malone@jiffyshirts.com