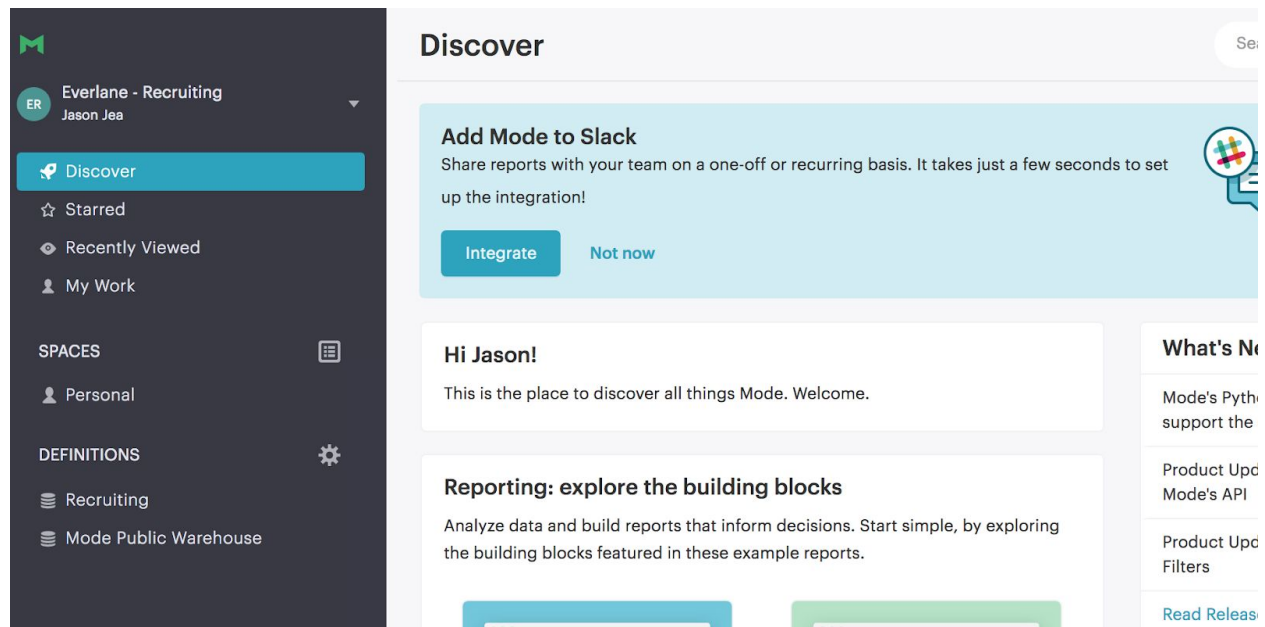
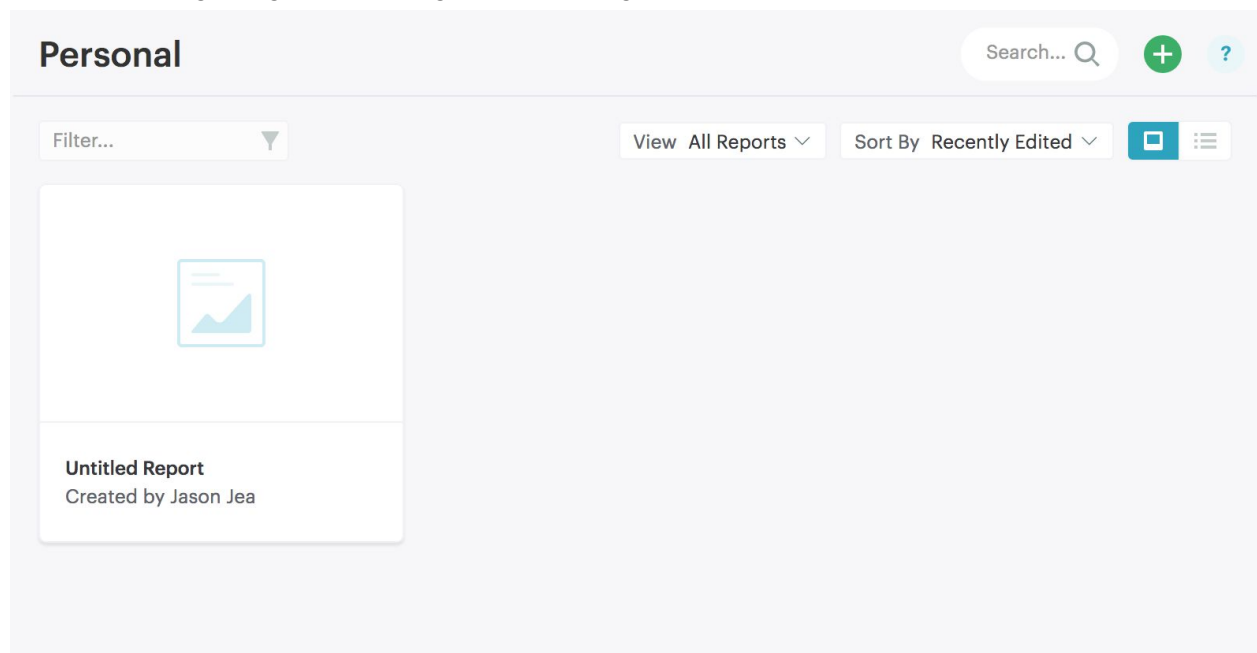


Instructions:

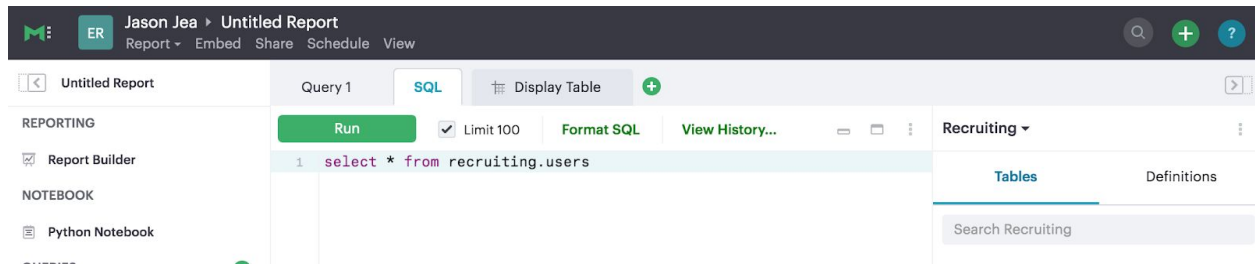
I've sent you an invite to Mode Analytics- please follow instructions to create an account. Once done, you can create a new workbook by navigating to your **Personal** space on the left hand toolbar:



And then hitting the green plus sign in the top right corner to create a new notebook:



Once your notebook is created, you can bookmark the URL to always navigate back to it. Next, click the drop down arrow on the right hand side (it might say **Mode Public Warehouse** at the moment), and choose **Recruiting** instead.



You're all set to start playing with some data!

If you're unfamiliar with SQL no worries- the datasets are small enough that you can pull into R, Python, or your analysis tool of choice. Just uncheck the Limit 100 box, run a simple select * statement on a table, then hit the **Export** button! Additionally, you can easily combine SQL with another analysis tool by following the steps above with a more complex query.

Tips:

The entire project should take between 3-5 hours. The most successful projects do not overcomplicate approaches and are specific in their answers. We understand there are many many ways to look at the data and answer the questions- try to avoid getting caught in analysis paralysis and spending time digging into every possible hypothesis. Instead, focus on fleshing out one or two approaches that you have committed to. We *will* occasionally ask "future analysis" type questions, which is your time to verbally lay out potential other ways to approach the dataset.

When returning your project, it's helpful to provide comments or instructions that allow us to follow your analysis workflow (i.e. I ran this SQL statement, then downloaded the csv, then loaded into R for analysis).

For this project, you will need to use these three tables:

- **recruiting.users**
 - A discovery is the first time we see a user on everlane.com.
 - We bucket marketing channels into five primary categories: paid (which can be broken into smaller channels), direct (arriving at everlane by typing in everlane.com directly), email, organic search (clicking through a non-ad result on a search engine), and other
 - Within paid, there are a few more channel designations: social (Facebook/Pinterest/Instagram ads), display (banners or sponsored article links on other websites), search (paid ads on search engines), affiliate (bloggers and influencers), and other
- **recruiting.orders**
 - This table is at the line item level- you can have multiple line items in one distinct order. A line item will map to a distinct style.
 - It also contains information on the channel that drove the order, with the same designations as you'll find in recruiting.users
- **recruiting.products**
 - This table contains basic information for our products.

Background:

You are tasked with providing some recommendations to Everlane around product launch strategy and customer acquisition. Everlane likes to look at its business through two lenses: the weekly new product launches that drive a big part of its revenue, and also the number of new customers making their first purchase.

Please return your answers with all code, visuals, and tables that you used to complete the project. The more information you can give us about your thought process, the better we can evaluate your project.

Questions:

1. Customer segments:

- a. Our marketing team is interested in creating a new email campaign focused on "Shoe (Footwear) Lovers". Using the data at hand, what is a way that we can identify shoe lovers?
- b. Let's say we wanted to generalize this campaign to every category in our product catalog (i.e. Shoe Lovers, Sweater Lovers, etc.). Please write code that will generate three lists of user_ids: one of Shoe (Footwear) Lovers, one of Sweater Lovers, and one of Knit Top Lovers. Bonus points if you can generalize this across all categories.

- c. Are there any conclusions you can draw about the customer profile of Shoe Lovers vs. Knit Top Lovers?

2. Experimentation:

- a. In order to measure the success of this new Shoe Lover email campaign, we would like to create an A/B test. How would you design an experiment to measure the impact of such a campaign? Please include an explanation for how you would collect and store data in order to measure such impact.

3. Cookie pairing:

- a. Imagine a dataset with the following format:

ip_address	cookie_id	user_id
1.1.1	abc	123
1.1.1	cbd	
2.2.2	abc	123
2.2.2	def	123
2.2.2	ghi	356
3.3.3	jkl	

A **cookie_id** represents a distinct device.

A **user_id** will represent an authenticated user.

- Multiple **cookie_id**'s can belong to the same **user_id** (these are people logging in on different devices).
- Assume that only one **user_id** max can be associated with a **cookie_id** (in other words, you will not see multiple **user_id**'s belonging to one **cookie_id**). The assumption here is that one device will only ever have one **user_id** authenticated with it.
- A **user_id** can show up on an infinite amount of **ip_addresses**, as can a **cookie_id**.

Let's say we are interested in understanding if a pair of **cookie_id**'s on the same **ip_address** belong to the same **user_id**. Step one would be to build a dataset that looks at every pairwise combination of **cookie_id**'s on the same **ip_address**, and label each pair as belonging to the same user or not. We can exclude **ip_address**'s that only have one **cookie_id**.

Please write code in your language of choice to produce the dataset described above. The end result should be something like:

ip_address	cookie_id_1	cookie_id_2	is_same_user
1.1.1	abc	cbd	0
2.2.2	abc	def	1
2.2.2	abc	ghi	0
2.2.2	def	ghi	0