

Terner Labs Economist/Data Analyst Technical Exercise

This technical exercise has two components and is meant to be completed in around 2-3 hours.

In the first section we are testing for basic coding competency in Python and use of geopandas. We want to see you join some datasets and generate some insights. Please document your code. We want to understand how you approach problems, so feel free to provide some narration at moments where you make a judgment call. If you are unsure of something, just state your assumptions and keep going.

If you have any questions or feel that you cannot move forward without a clarifying question, please don't hesitate to reach out to alex.casey@ternerlabs.org, darrell.owens@ternerlabs.org and haley.tiu@ternerlabs.org. Please submit a single zip file that contains all of your work, as well as all data needed to run your code.

Data
City of San Jose Parcel Level Data
City of San Jose Zoning Districts
Census Data: For section A questions 3 & 4 you are welcome to use any form of census data you choose, whether that be through the API, a package you prefer, the census website , or IPUMS microdata. Please be sure to document any choices you make in selecting data.

Section A:

1. Find the San Jose Parcels in Urban Village, Mixed Use Commercial, Urban Residential, and Transit Residential zones (zoning codes [here](#)). How many Parcels are there in each of these zones?

- Join the zoning geometries to the parcel geometries to find which zone the parcel falls into
- Note that some parcels may overlap with more than one zone. We suggest you write a custom function in a separate .py file that returns only one overlapping zone per parcel
- Determine how many parcels are in each zone

2. Calculate the total acreage across all parcels in each zone. Which zone, Urban Village or Urban Residential has the highest possible number of dwelling units based solely on the maximum dwelling unit per acre (du/ac) allowance for 100% residential development? You will need table Table 20-136 in the San Jose city code [here](#). Also, please note the unit for the SHAPE_Area variable in the Parcel data set is in square feet.

3. Calculate the number of Urban Village parcels within a mile of the San Jose Diridon Station.
4. Choose two or three census characteristics related to evaluating aspects of opportunity, economic mobility, or environmental sustainability near transit stations that might inform policy design. Provide a 1-2 paragraph summary and at least one visualization that you think a San Jose government official might want to see to better understand the characteristics of the area within two miles of Diridon Station. As indicated in the data section above you can use whatever source of census data you are most comfortable with. You may need to perform a geographic crosswalk here, which you may do however you wish.

Section B:

This open-ended section provides an opportunity to showcase data skills relevant to the role. This is meant to be relatively brief - a paragraph or two describing the project and then around 40 lines¹ of code in Python. For a non-exhaustive list of possibilities: Is there a tricky section of a data pipeline that highlights your meticulous nature? Did you develop a useful predictive model? Do you have a data visualization that you think is impactful? Do you have a particularly thoughtful way of doing exploratory data analysis that you want to show off? Sharing code via GitHub is preferred but not required.

¹ This is just to provide some general guidelines, there is certainly no penalty for going under or over these numbers. The quality of writing and coding is, of course, not related to its length.