

2B. Evidence of Discrimination__

October 10, 2021

1 Evidence of Discrimination?

The Department of Developmental Services (DDS) in California is responsible for allocating funds to support over 250,000 developmentally-disabled residents. The data set `ca_dds_expenditures.csv` contains data about 1,000 of these residents. The data comes from a discrimination lawsuit which alleged that California's Department of Developmental Services (DDS) privileged white (non-Hispanic) residents over Hispanic residents in allocating funds. We will focus on comparing the allocation of funds (i.e., expenditures) for these two ethnicities only, although there are other ethnicities in this data set.

There are 6 variables in this data set:

- Id: 5-digit, unique identification code for each consumer (similar to a social security number and used for identification purposes)
- Age Cohort: Binned age variable represented as six age cohorts (0-5, 6-12, 13-17, 18-21, 22-50, and 51+)
- Age: Unbinned age variable
- Gender: Male or Female
- Expenditures: Dollar amount of annual expenditures spent on each consumer
- Ethnicity: Eight ethnic groups (American Indian, Asian, Black, Hispanic, Multi-race, Native Hawaiian, Other, and White non-Hispanic)

2 Question 1

Read in the data set. Make a graphic that compares the *average* expenditures by the DDS on Hispanic residents and white (non-Hispanic) residents. Comment on what you see.

```
[1]: # YOUR CODE HERE
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

ca_dds_data = pd.read_csv("ca_dds_expenditures.csv")
hispanicData = ca_dds_data.loc[ca_dds_data["Ethnicity"] == "Hispanic"]
whiteData = ca_dds_data.loc[ca_dds_data["Ethnicity"] == "White not Hispanic"]
comparionData = whiteData.append(hispanicData)
```

```

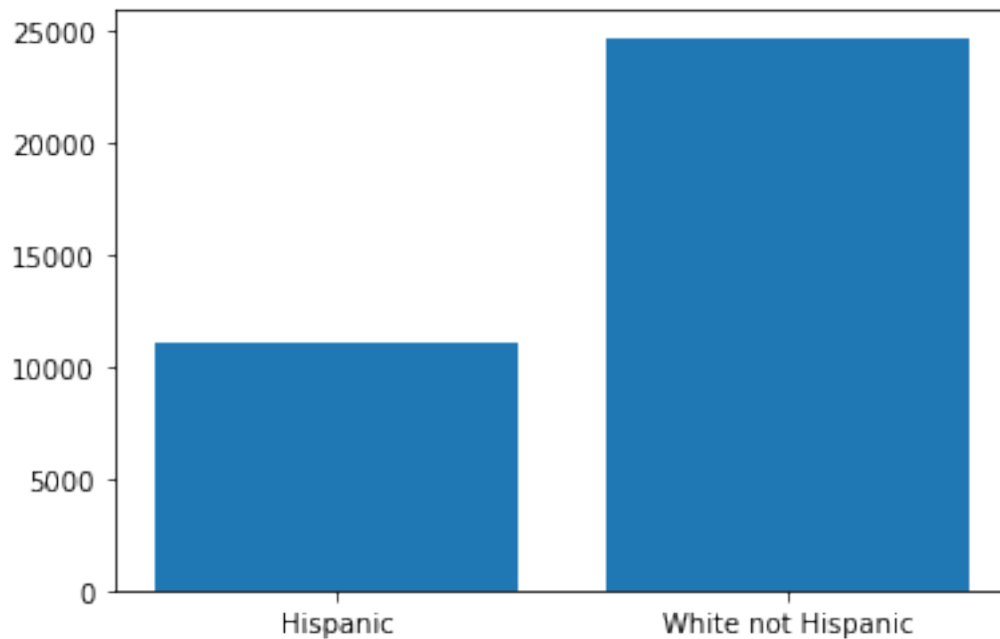
ethnicity = ["Hispanic", "White not Hispanic"]
data = comparisonData.groupby("Ethnicity")["Expenditures"].mean()

plt.bar(ethnicity, data)
plt.plot()

# ca_ddc_data.groupby("Ethnicity").mean()

```

[1]: []



YOUR EXPLANATION HERE The white not hispanic has around 2.5 times more expenditures than the hispanic so therefore it seems like there is some discrimination.

3 Question 2

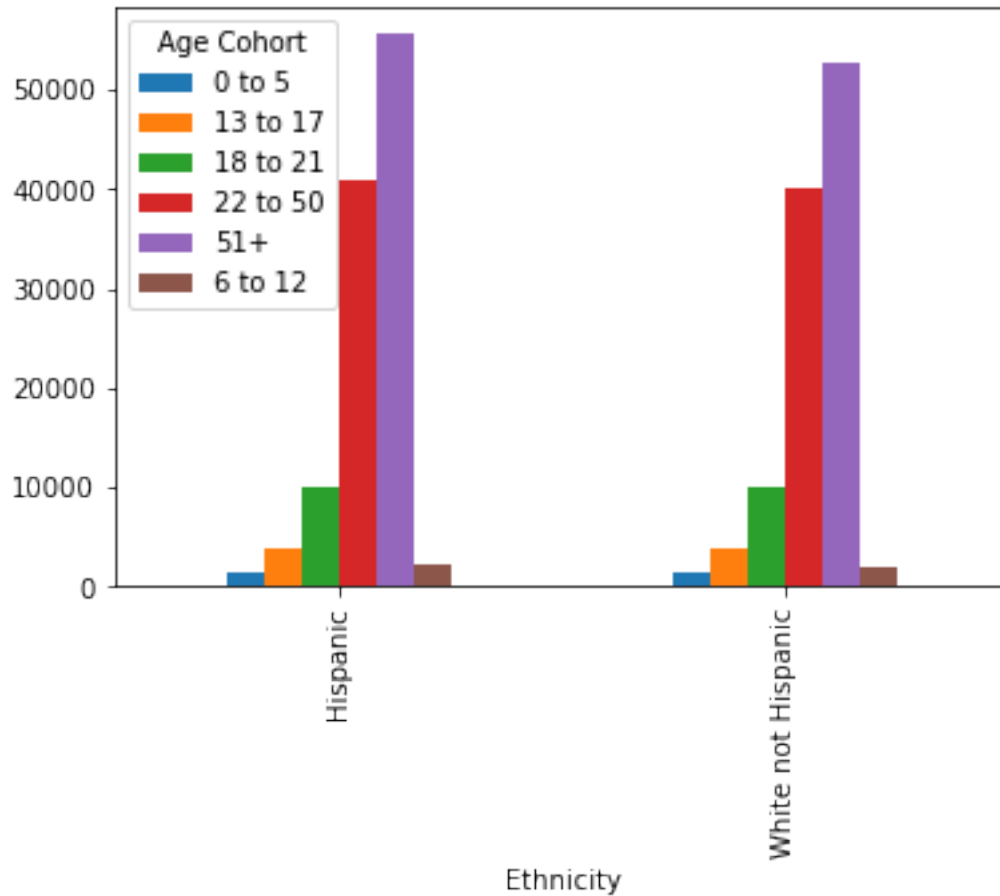
Now, calculate the average expenditures by ethnicity and age cohort. Make a graphic that compares the average expenditure on Hispanic residents and white (non-Hispanic) residents, *within each age cohort*.

Comment on what you see. How do these results appear to contradict the results you obtained in Question 1?

[2]: # YOUR CODE HERE

```
age_cohortGraph = ca_dds_data.pivot_table(index="Ethnicity", columns="Age_Cohort", values="Expenditures")
age_cohortGraph.loc[["Hispanic", "White not Hispanic"]].plot.bar()
```

[2]: <AxesSubplot:xlabel='Ethnicity'>



YOUR EXPLANATION HERE This contradicts question 1 because it seems like there is no discrimination. This graph also shows that the older you are the more money you will spend.

4 Question 3

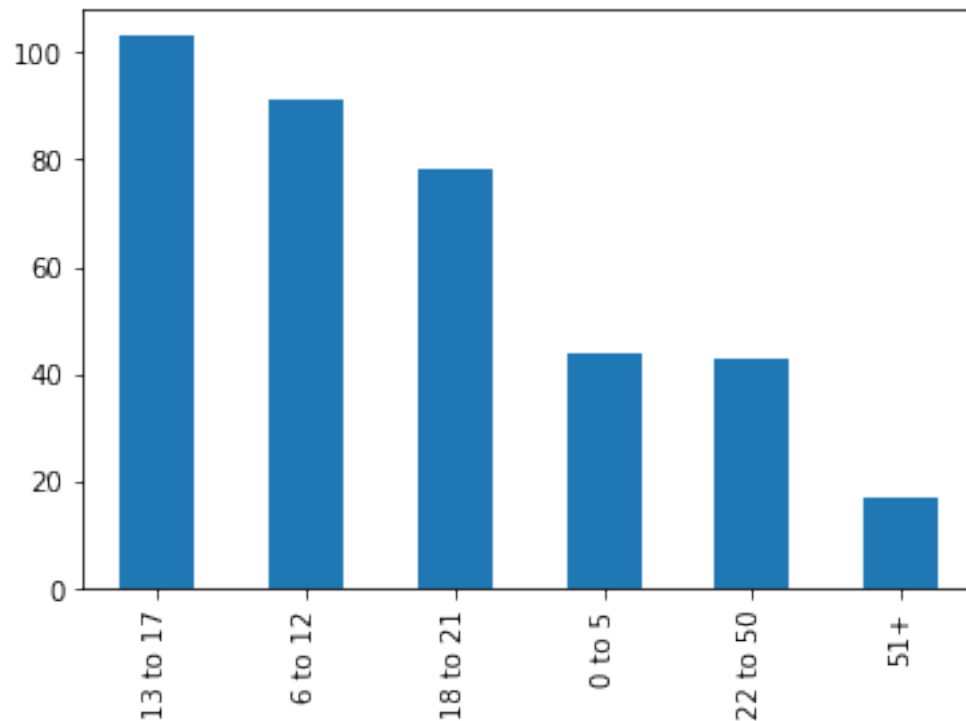
Can you explain the discrepancy between the two analyses you conducted above (i.e., Questions 1 and 2)? Try to tell a complete story that interweaves tables, graphics, and explanation.

Hint: You might want to consider looking at:

- the distributions of ages of Hispanics and whites
- the average expenditure as a function of age

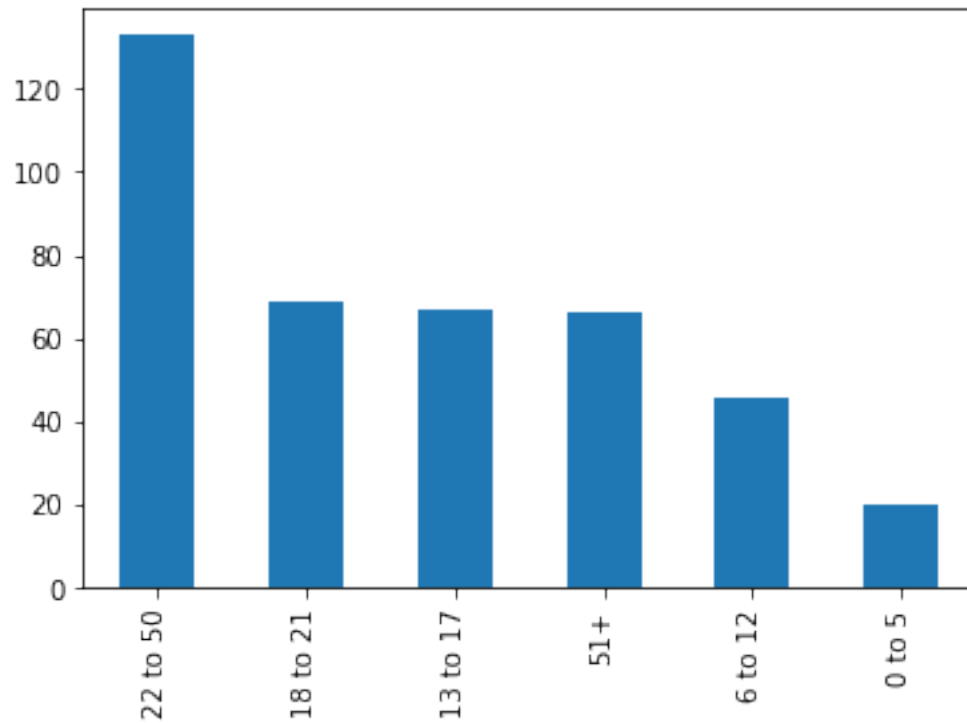
```
[3]: # YOUR CODE HERE (although you may want to add more code cells)
ca_dds_data.loc[ca_dds_data["Ethnicity"] == "Hispanic", "Age Cohort"].
    ↳value_counts().plot.bar()
```

[3]: <AxesSubplot:>



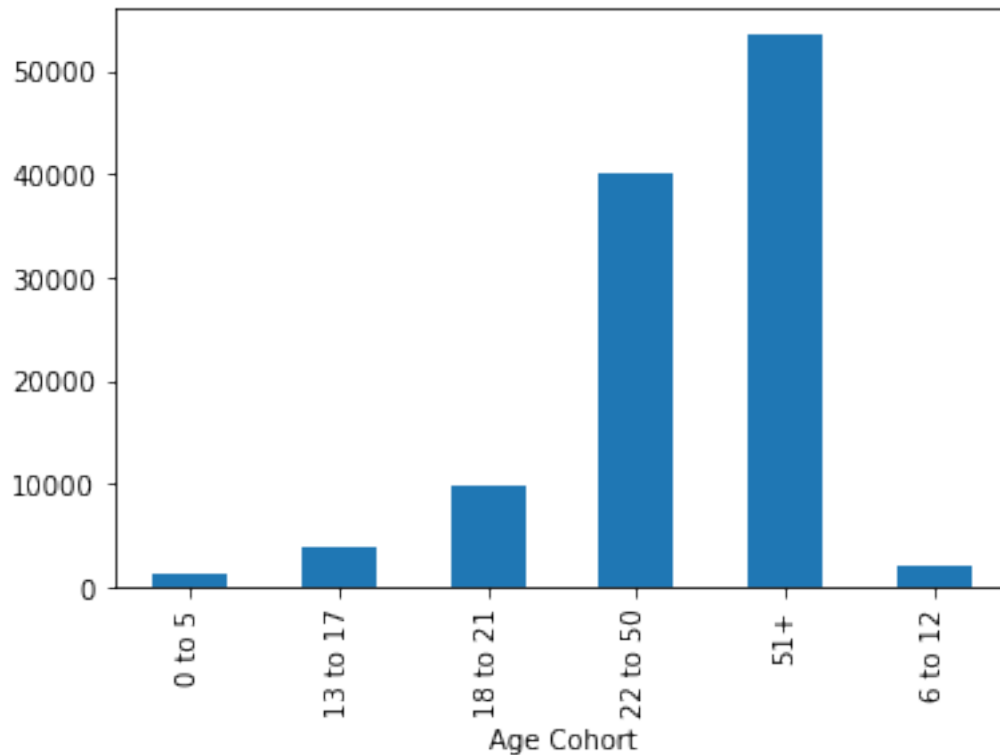
```
[4]: ca_dds_data.loc[ca_dds_data["Ethnicity"] == "White not Hispanic", "Age Cohort"].
    ↳value_counts().plot.bar()
```

[4]: <AxesSubplot:>



```
[5]: ca_dds_data.groupby("Age Cohort")["Expenditures"].mean().plot.bar()
```

```
[5]: <AxesSubplot:xlabel='Age Cohort'>
```



YOUR EXPLANATION HERE (although you may want to add more markdown cells)

There are couple of important details in these graphs:

- There 66 White non Hispanic in the age cohort 51+ and only 17 Hispanic in the age cohort 51+
- The Hispanic population is very young while White not Hispanic is much older. Since there is positive correlation between age and expenditure, the graph from question 1 is skewed because Hispanics are younger.

4.1 Submission Instructions

Once you are finished, follow these steps:

1. Restart the kernel and re-run this notebook from beginning to end by going to **Kernel > Restart Kernel and Run All Cells**.
2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end.
3. Double check that there is a number next to each code cell and that these numbers are in order.

Then, submit your lab as follows:

1. Go to **File > Export Notebook As > PDF**.

2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.)
3. Upload the PDF to Gradescope.
4. Demo your lab.