

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)

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.

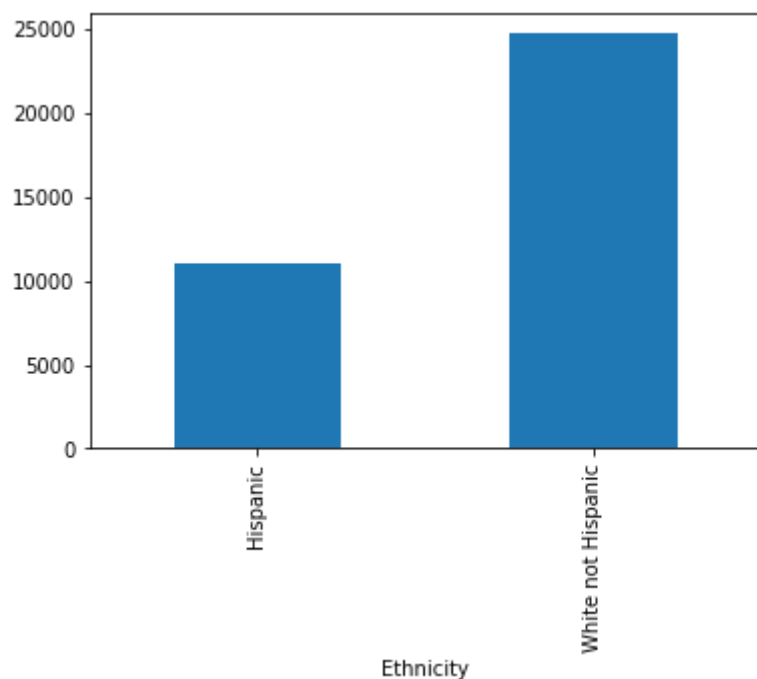
In [1]:

```
# YOUR CODE HERE
import pandas as pd
df = pd.read_csv("ca_dds_expenditures.csv")

col1 = df.loc[df.Ethnicity == "Hispanic"]
col2 = df.loc[df.Ethnicity == "White not Hispanic"]
col1 = col1.append(col2)
col1.groupby("Ethnicity")["Expenditures"].mean().plot.bar()
```

Out[1]:

<matplotlib.axes._subplots.AxesSubplot at 0x7fa5fb7cc610>



YOUR EXPLANATION HERE We can see that on average, White not Hispanic people have a higher average expenditure according to the DDS.

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*.

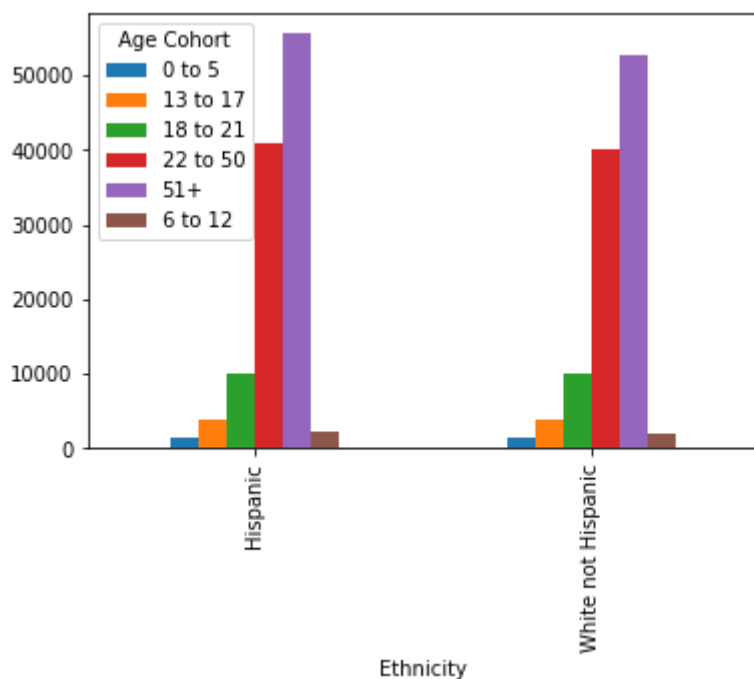
Comment on what you see. How do these results appear to contradict the results you obtained in Question 1?

In [2]:

```
# YOUR CODE HERE
import numpy as np
AvgExp = df.pivot_table(index = "Ethnicity", columns = "Age Cohort", values = "Expenditures", aggfunc = np.mean)
AvgExp.loc[["Hispanic", "White not Hispanic"]].plot.bar()
```

Out[2]:

<matplotlib.axes._subplots.AxesSubplot at 0x7fa5fb94ead0>



YOUR EXPLANATION HERE The distributions for Hispanic and White across all age groups are pretty similar. It seems ages 51+ have the highest average expenditure. This contradicts what we saw in Question 1 because we saw that the average expenditure is higher for Whites.

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:

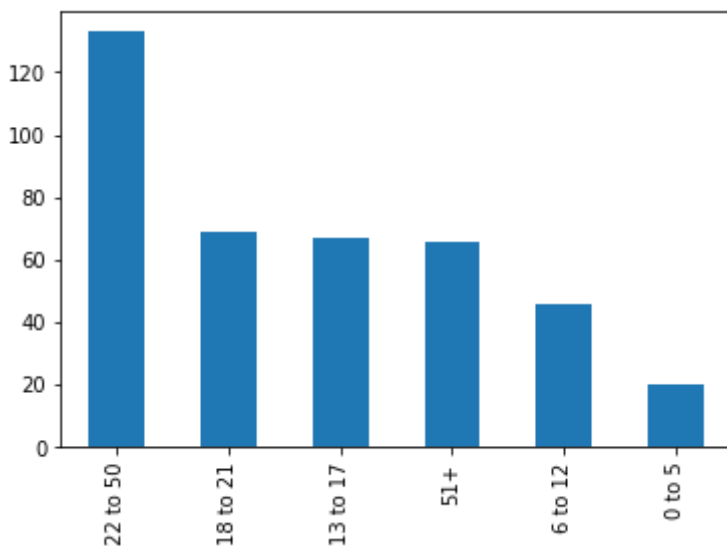
- the distributions of ages of Hispanics and whites
- the average expenditure as a function of age

In [3]:

```
# YOUR CODE HERE (although you may want to add more code cells)
import pandas as pd
df = pd.read_csv("ca_dds_expenditures.csv")
bar_exp = df.loc[df["Ethnicity"] == "White not Hispanic", "Age Cohort"].value_counts()
bar_exp.plot.bar()
bar_exp
```

Out[3]:

```
22 to 50    133
18 to 21     69
13 to 17     67
51+          66
6 to 12      46
0 to 5       20
Name: Age Cohort, dtype: int64
```



YOUR EXPLANATION HERE (although you may want to add more markdown cells) The discrepancy between 1 and 2 is that for 2 we are looking at a more specific subset of Average expenditure (by age). When we compare Whites and Hispanics overall it appears that Hispanics have a lower average expenditure but when analyzing the two groups individually we can see that the distribution across age groups is very similar, even though on AVERAGE, Whites have a higher expenditure.

In []:

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.

In []: