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:

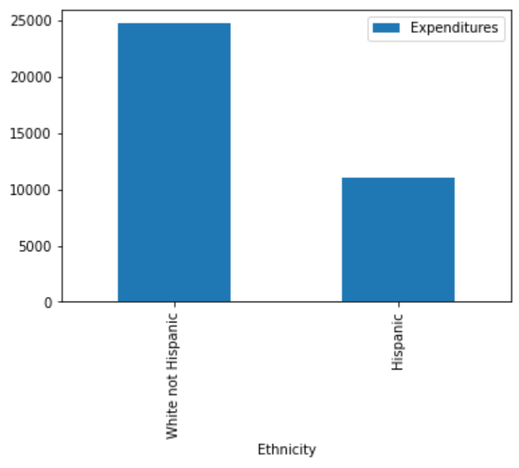
- Id: 5-digit, unique identification code for each consumer (similar to a social security number and used for identification purposes)
- Age Cohort: Binned age variable represented as six age cohorts (0-5, 6-12, 13-17, 18-21, 22-50, and 51+)
- Age: Unbinned age variable
- Gender: Male or Female
- Expenditures: 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
import numpy as np
df = pd.read_csv("ca_dds_expenditures.csv")
pivot_table = df.pivot_table(index = "Ethnicity", values = "Expenditures", aggfunc = np.mean)
pivot_table.loc[["White not Hispanic","Hispanic"]].plot.bar()
```

Out[1]: <AxesSubplot:xlabel='Ethnicity'>



YOUR EXPLANATION HERE There does appear to be a noticable increase between the average money given to white (not hispanic) people in DDS compared to hispanic people.

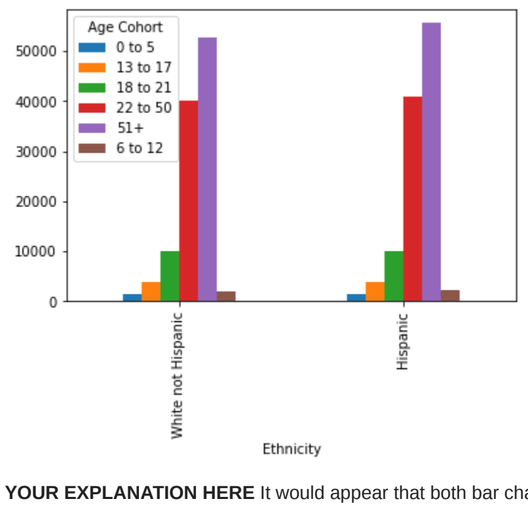
Question 2

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
pivot_table = df.pivot_table(index = "Ethnicity", values = "Expenditures", columns = "Age Cohort", aggfunc = np.mean)
pivot_table.loc[["White not Hispanic","Hispanic"]].plot.bar()
```

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



YOUR EXPLANATION HERE It would appear that both bar charts appear to show very similar trends of spending. Which is contradictory from the result in Q1.

Question 3

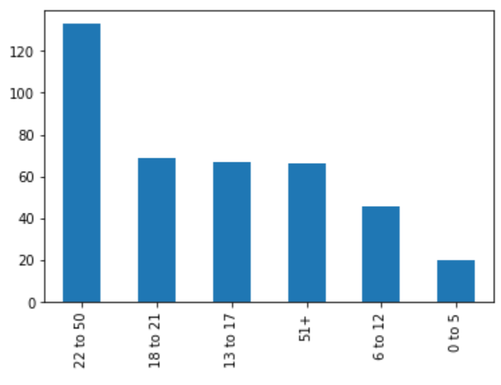
Can you explain the discrepancy between the two analyses you conducted above (i.e., Questions 1 and 2)? Try to tell a complete story that interweaves tables, graphics, and explanation.

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)
df.loc[df["Ethnicity"] == "White not Hispanic", "Age Cohort"].value_counts().plot.bar()
```

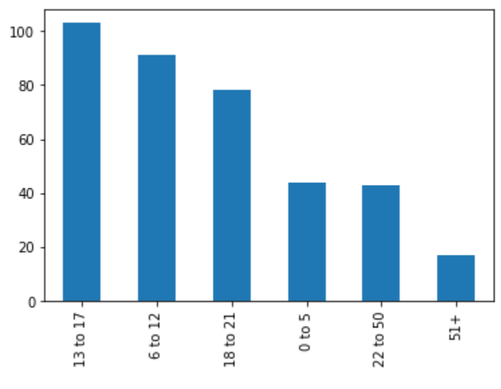
Out[3]: <AxesSubplot:>



YOUR EXPLANATION HERE The distribution of ages for White not Hispanic people. 22 to 50 is the largest category of people receiving aid.

In [4]:
```
# YOUR CODE HERE (although you may want to add more code cells)
df.loc[df["Ethnicity"] == "Hispanic", "Age Cohort"].value_counts().plot.bar()
```

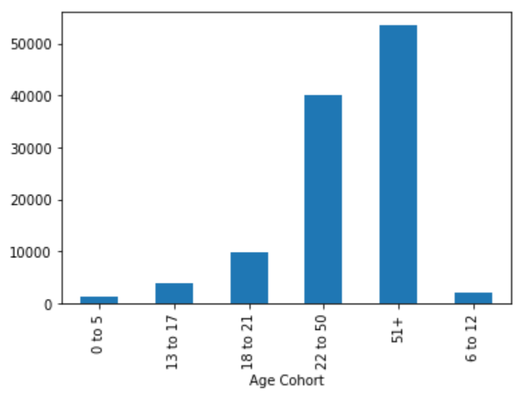
Out[4]: <AxesSubplot:>



YOUR EXPLANATION HERE The distribution of ages for Hispanic people. 13 to 17 is the largest category of people receiving aid.

In [5]:
```
# YOUR CODE HERE (although you may want to add more code cells)
meanPerAgeRange = df.groupby("Age Cohort")["Expenditures"].mean()
meanPerAgeRange.plot.bar()
```

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



YOUR EXPLANATION HERE It seems that the age group of 51+ receives the most average amt. of money, regardless of race.

Conclusion: Connecting to why White people receive more average amt of money than Hispanic people, because there are more 51+ year old white people than Hispanic people, and 51+ year old people are the ones to receive the most money, regardless to age, there should be no surprise that white people receive a larger amount of money of DDC.

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 $Ker \neq l > Re \star tKer \neq l$ and $RunAllCells$. 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 $Fi \leq > Exp$ or $tNoteb \infty kAs > PDF$. 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.