

# Chatlize generated report

Chatlize.ai

2023-11-13

## System

Act as an experienced data scientist and statistician. Write Python code to analyze data following instructions.

## Step 1

Prompt: Write Python code. Do not include any explanation. Read the file 'data\_earnings.csv' in the current folder and print a small part of it. These columns are numeric variables: 'age', 'education', 'earnings', 'hours', and 'id'. These columns are categorical variables: 'sex', and 'edcat'.

```
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv('data_earnings.csv')

# Print a small part of the DataFrame to inspect it
print(df.head())
```

```
##      sex  age  education  earnings  hours  id  edcat
## 0  male   40         9.0   22000.0  2340.0  1.0    1.0
## 1  male   37        12.0   31468.0  2008.0  2.0    2.0
## 2  male   43        16.0   38708.0  3240.0  3.0    4.0
## 3  male   32        12.0   40000.0  2200.0  4.0    2.0
## 4  male   48         9.0         0.0    0.0  5.0    1.0
```

## Step 2

Prompt: Write Python code. Do not include any explanation. drop 0 and absent values from data

```
# Drop rows with 0 or NaN (absent) values
df = df.replace(0, pd.NA)
df = df.dropna()

# Print the updated DataFrame to verify changes
print(df.head())
```

```
##      sex  age  education  earnings  hours  id  edcat
## 0  male   40         9.0   22000.0  2340.0  1.0    1.0
## 1  male   37        12.0   31468.0  2008.0  2.0    2.0
## 2  male   43        16.0   38708.0  3240.0  3.0    4.0
## 3  male   32        12.0   40000.0  2200.0  4.0    2.0
## 5  male   30        12.0    3265.0   552.0  6.0    2.0
```

## Step 3

Prompt: Write Python code. Do not include any explanation. show data (20 first rows)

```
# Show the first 20 rows of the DataFrame
print(df.head(20))
```

```
##      sex  age  education  earnings   hours   id  edcat
## 0   male   40         9.0   22000.0   2340.0   1.0    1.0
## 1   male   37        12.0   31468.0   2008.0   2.0    2.0
## 2   male   43        16.0   38708.0   3240.0   3.0    4.0
## 3   male   32        12.0   40000.0   2200.0   4.0    2.0
## 5   male   30        12.0    3265.0    552.0   6.0    2.0
## 6   male   45        11.0   18000.0   2286.0   7.0    1.0
## 7   male   37        17.0   71750.0   3750.0   8.0    4.0
## 8   male   47        11.0   19426.0   2016.0   9.0    1.0
## 9   male   38        12.0    4000.0   2288.0  10.0    2.0
## 11  male   33        11.0   30000.0   2860.0  12.0    1.0
## 12  male   45        13.0   24000.0   2820.0  13.0    3.0
## 13  male   30        15.0   14400.0   2290.0  14.0    3.0
## 14  male   39        12.0   30000.0   2000.0  15.0    2.0
## 15  male   43        12.0   69000.0   2100.0  16.0    2.0
## 16  male   38        12.0   50430.0   1838.0  17.0    2.0
## 17  male   30        17.0   30000.0   1720.0  18.0    4.0
## 18  male   34        12.0   34000.0   2744.0  19.0    2.0
## 19  male   33        12.0   22000.0   2304.0  20.0    2.0
## 20  male   35        12.0   50300.0   2503.0  21.0    2.0
## 21  male   38        12.0   32629.0   2400.0  22.0    2.0
```

## Step 4

Prompt: Write Python code. Do not include any explanation. make bar chart using data where x equals to edcat and y equals average earnings, add axis labels, heading and coloring by edcat

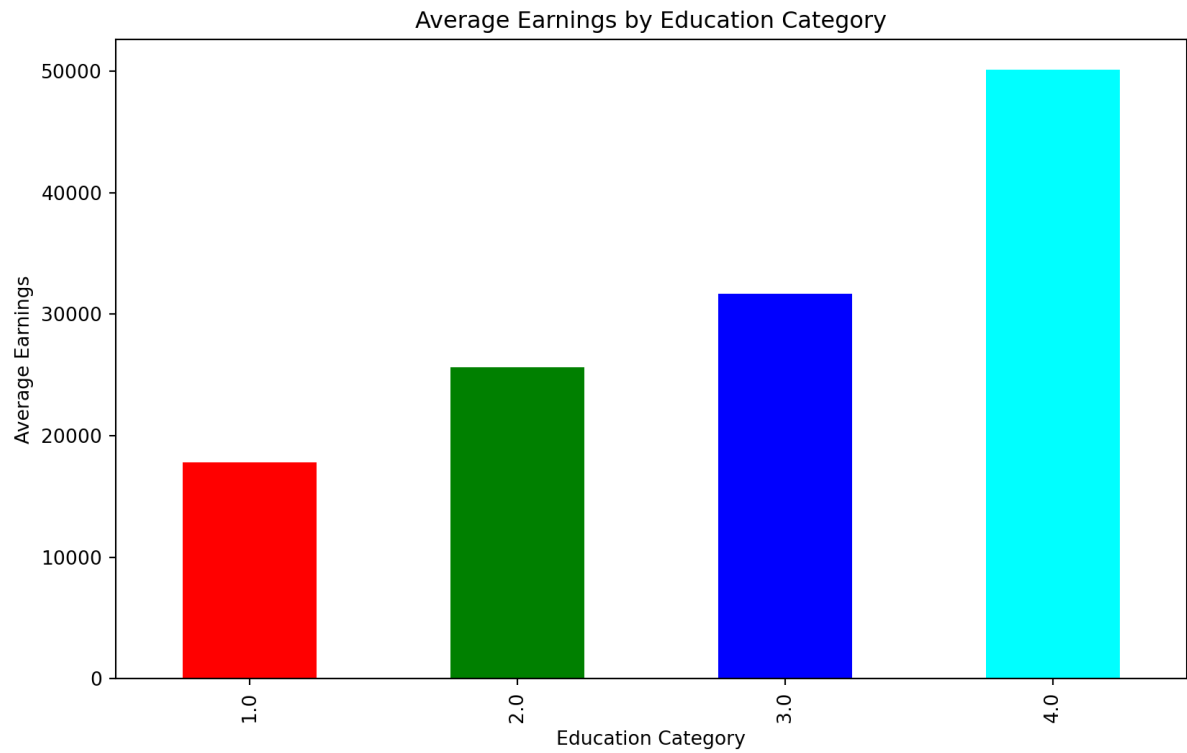
```
import matplotlib.pyplot as plt

# Group the data by 'edcat' and calculate the average earnings for each category
average_earnings_by_edcat = df.groupby('edcat')['earnings'].mean()

# Create the bar chart
average_earnings_by_edcat.plot(kind='bar', color=['red', 'green', 'blue', 'cyan', 'magenta'])

# Add labels and title
plt.xlabel('Education Category')
plt.ylabel('Average Earnings')
plt.title('Average Earnings by Education Category')

# Display the bar chart
plt.show()
```



Generated with <https://Chatlize.ai> (<https://Chatlize.ai>). No guarantees! Still a work in progress. Contact Steven Ge ([gexijin@gmail.com](mailto:gexijin@gmail.com) (<mailto:gexijin@gmail.com>)).