



IBM assignment

Danai Avratoglou – BAFT1512

Katerina Kalaidopoulou – BAFT1509

Nikos Papakonstantinou – BAFT1509

Konstantinos Fraidakis – BAFT1507

Konstantinos Chronis – BAFT1502



- Watson Analytics contribution to data understanding
- Exploration example
- Prediction example
- Dashboard

Our team



Danai



Katerina



Nikos



Konstantinos F.



Konstantinos C.



Watson Analytics contribution to data understanding

1



Ask a question about your data

How to ask a question?

Census-Income
Mon, 2 at 3:49 PM



2

Starting points

Show Next >



What are the values of **capital gains** and **capital losses** by **country of birth** **father**?



What drives **instance weight**?



What is a predictive model for **num persons worked for employer**?



What is the trend of **weeks worked in year** over **age** by **sex**?



What are the values of **weeks worked in year** for each **education**?



What are the connections between **sex** and **citizenship**?

3

Create your own visualization

Comparison



Bar



Combination



Word cloud



Dial



Heatmap

Parts to whole



Treemap



Pie



Packed bubble

Trend and forecast



Line



Area



Combination

4

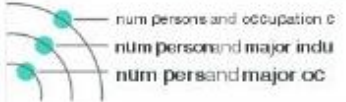


Discoveries

Top **age** by **weeks worked in year**



Top Drivers of **weeks worked in year**



40.4k

is the **weeks worked in year** for **citizenship**
Native- Born in Puerto Rico or U.S. Outlying

Top class of worker by **weeks worked in year**





Census Data*

39 Columns

146.141 Rows

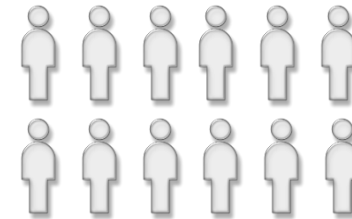
| age | sex | education | state of previ... | major industry ... | industry code | citizenship | class of worker | country of birth... | country of birth... | country of birth... | country of birth... | detailed house... | detailed house... | enrolled > |
|--------------------|--------------------|--------------------|-------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------------|--------------------|--------------------|--------------------|
| 75 High Quality | 96 High Quality | 58 Medium Quality | Unique values | 60 Medium Quality | 75 High Quality | 53 Medium Quality | 59 Medium Quality | 54 Medium Quality | 54 Medium Quality | 52 Medium Quality | | 62 Medium Quality | 60 Medium Quality | 52 Medium Quality |
| Missing Values: 0% | Missing Values: 0% | Missing Values: 0% | | Missing Values: 0% | Missing Values: 0% | Missing Values: 0% | Missing Values: 0% | Missing Values: 0% | Missing Values: 0% | Missing Values: 0% | | Missing Values: 0% | Missing Values: 0% | Missing Values: 0% |
| Number > | Text > | Text > | Text > | Text > | Number > | Text > | Text > | Text > | Text > | Text > | Hierarchy > | Text > | Text > | Text > |
| Average > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Count distinct > | Hierarchy > | Count distinct > | Count distinct > | Count distinct > |
| No sort > | No sort > | No sort > | No sort > | No sort > | No sort > | No sort > | No sort > | No sort > | No sort > | No sort > | Hierarchy > | No sort > | No sort > | No sort > |
| 73 | Female | High school gr... | Not in universe | Not in univers... | 0 | Native-Born i... | Not in universe | United-States | United-States | United-States | Not in universe | Other Rel 18+ ... | Other relative... | Not in univ |
| 58 | Male | Some college ... | Arkansas | Construction | 4 | Native-Born i... | Self-employe... | United-States | United-States | United-States | Arkansas | Householder | Householder | Not in univ |
| 18 | Female | 10th grade | Not in universe | Not in univers... | 0 | Foreign born... | Not in universe | Vietnam | Vietnam | Vietnam | Not in universe | Child 18+ neve... | Child 18 or old... | High schoi |
| 9 | Female | Children | Not in universe | Not in univers... | 0 | Native-Born i... | Not in universe | United-States | United-States | United-States | Not in universe | Child <18 neve... | Child under 1... | Not in univ |
| 10 | Female | Children | Not in universe | Not in univers... | 0 | Native-Born i... | Not in universe | United-States | United-States | United-States | Not in universe | Child <18 neve... | Child under 1... | Not in univ |
| 48 | Female | Some college ... | Not in universe | Entertainment | 40 | Native-Born i... | Private | Philippines | United-States | United-States | Not in universe | Spouse of hou... | Spouse of ho... | Not in univ |
| 42 | Male | Bachelors deg... | Not in universe | Finance insura... | 34 | Native-Born i... | Private | United-States | United-States | United-States | Not in universe | Householder | Householder | Not in univ |



1995

213109131.75

instance weight (Sum)

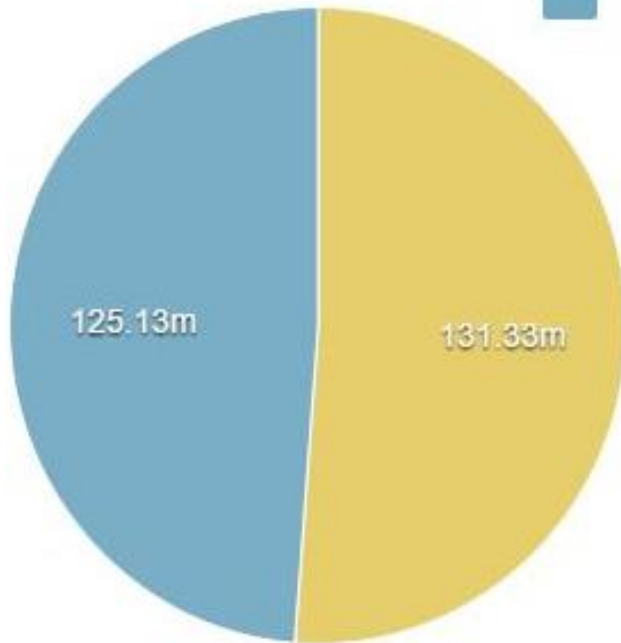


<http://archive.ics.uci.edu/ml/machine-learning-databases/census-income-mld/>

Data understanding

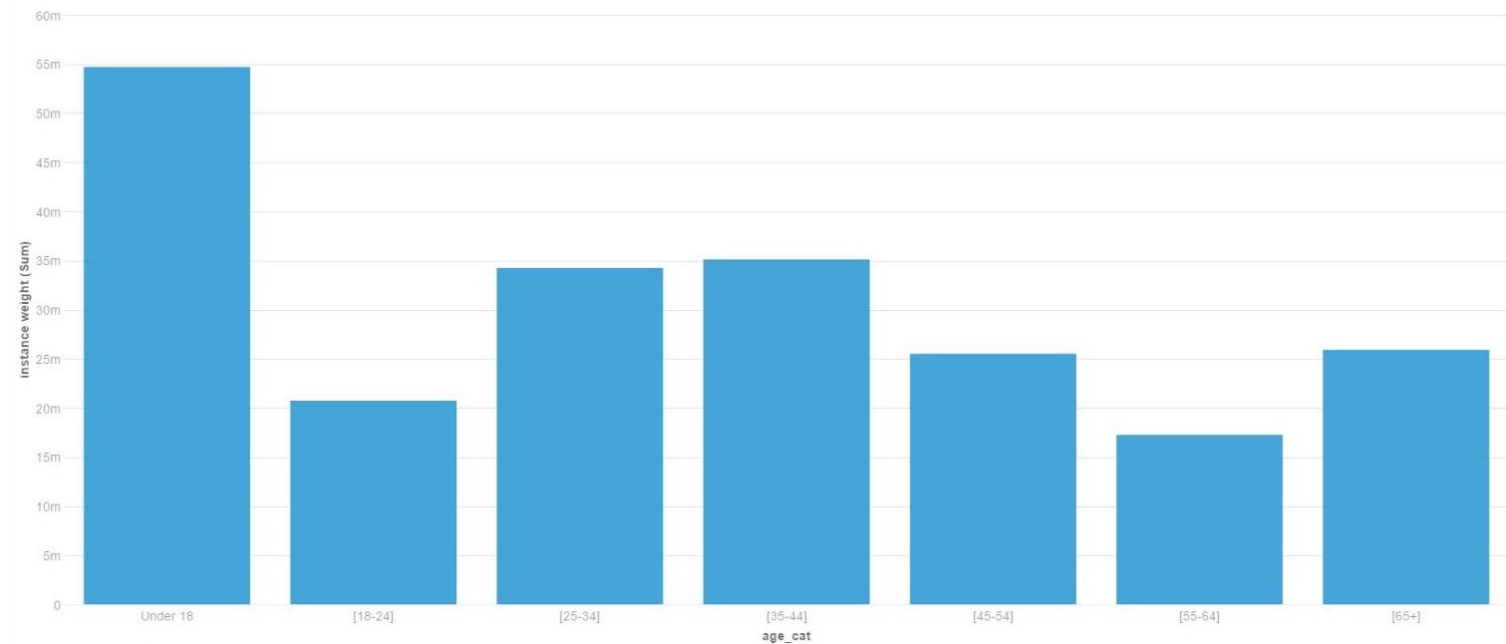
Distribution per

sex
Female
Male



Distribution per age group

How do the values of `instance weight` compare by `age_cat` ?



Data understanding

Top Education levels



Distribution per Marital status

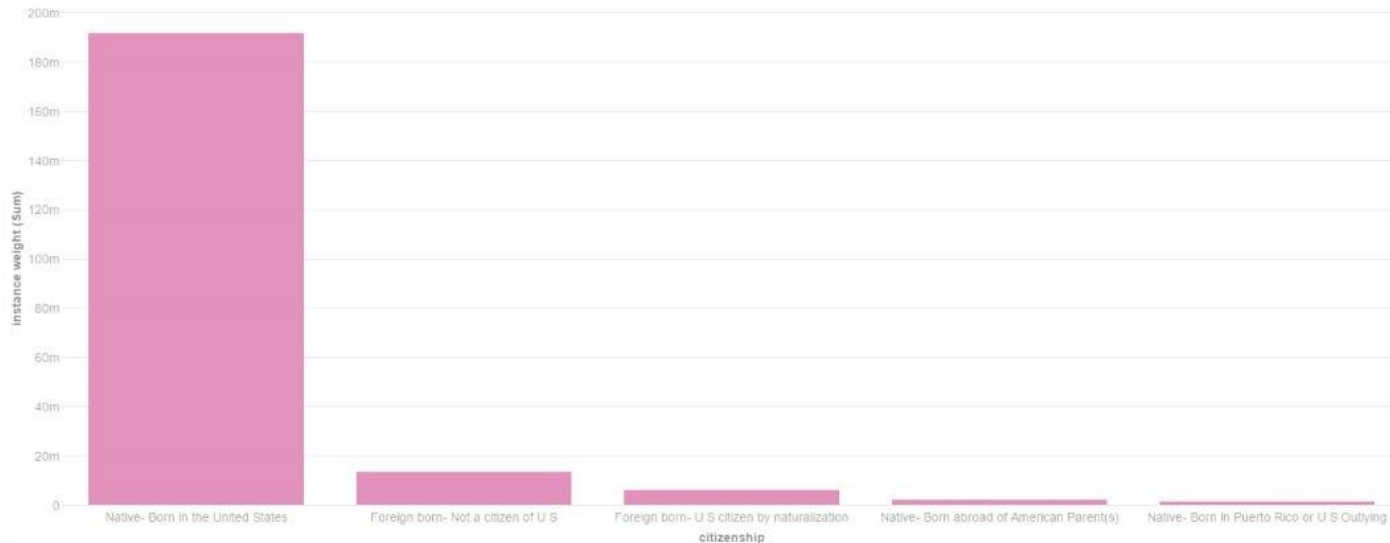
What is the breakdown of instance weight by marital status ?



Data understanding

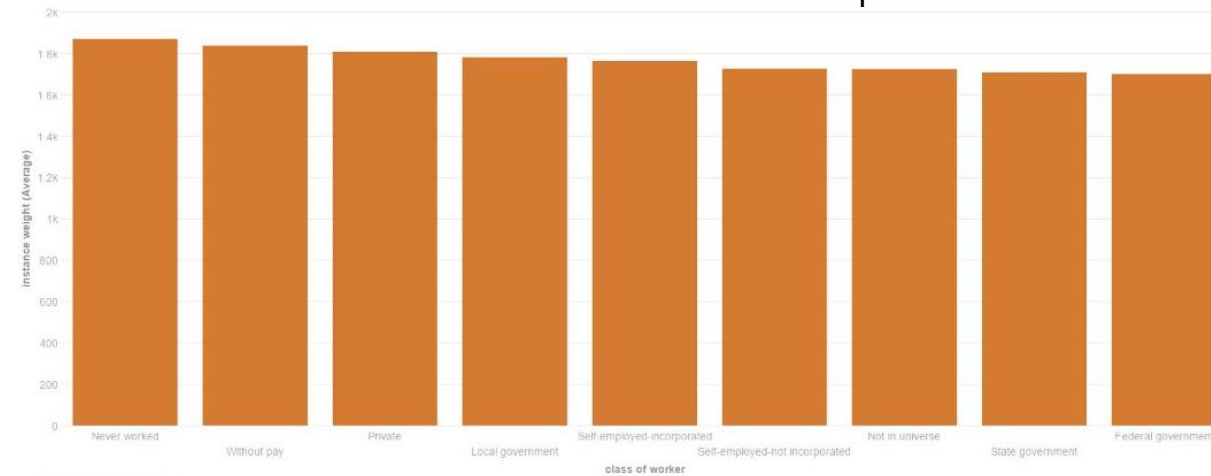
Distribution per Citizenship

How do the values of `instance weight` compare by `citizenship` ?

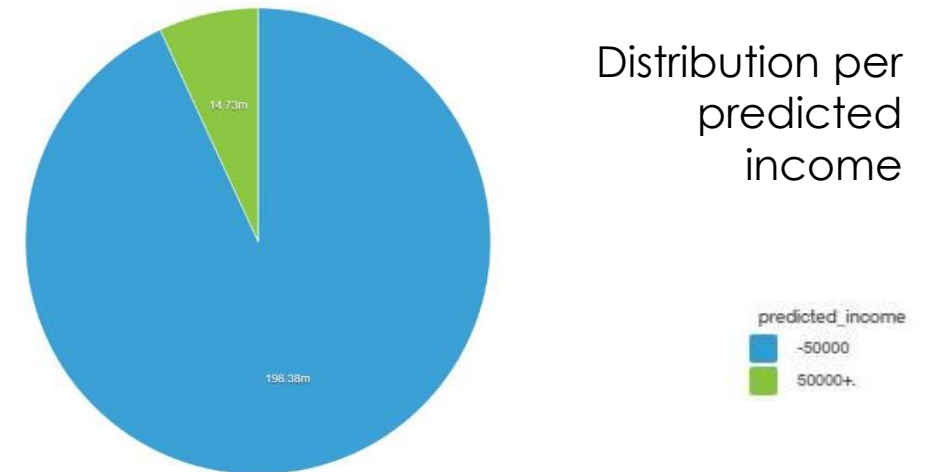


Distribution per class of worker

How do the values of `instance weight` compare by `class of worker` ?



What is the breakdown of `instance weight` by `predicted_income` ?



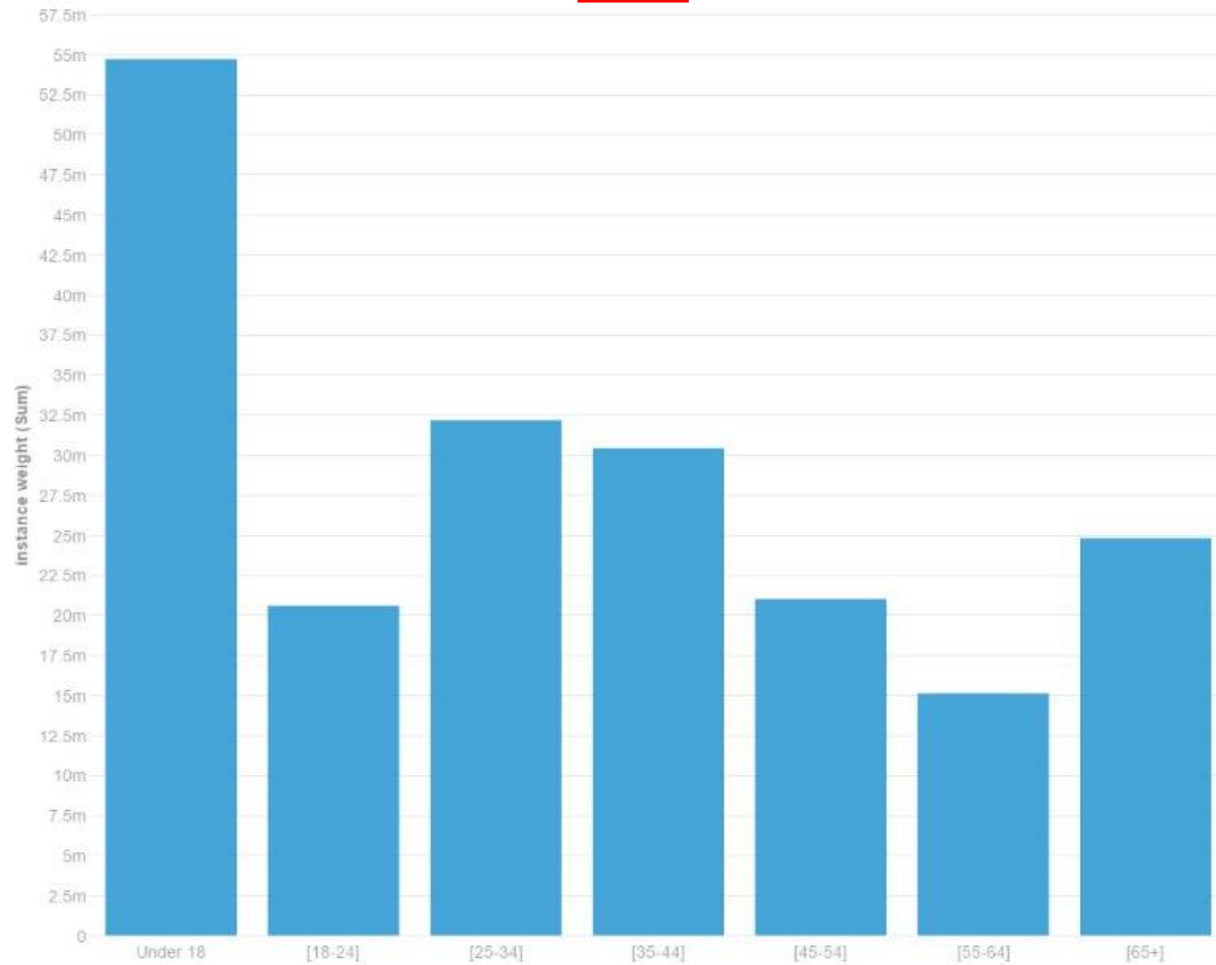
Distribution per
predicted
income

Income level per *age group*

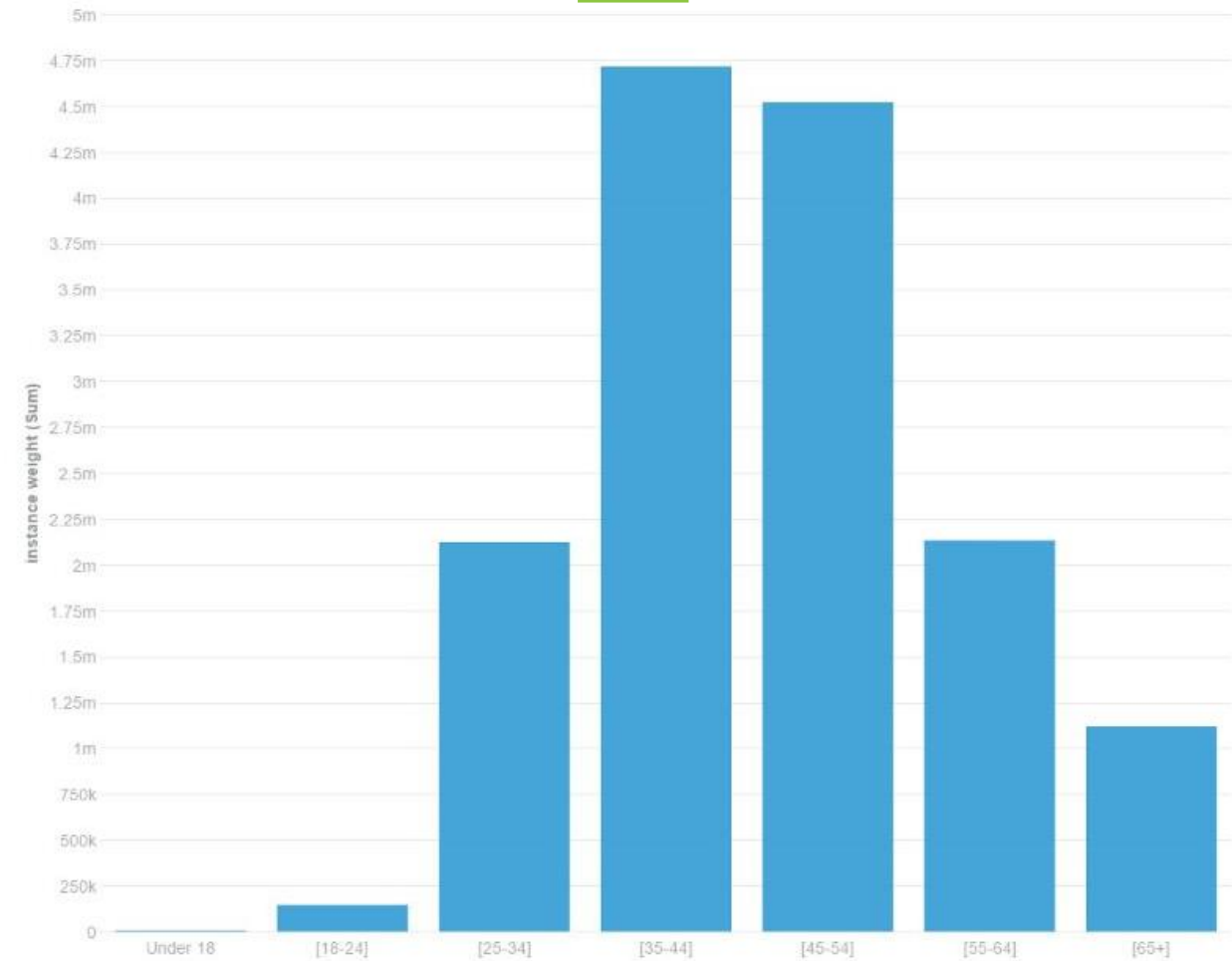
How do the values of **instance weight** compare by **age_cat** ?

Filtered by predicted_income: -50000, 50000+ (X)

-50000



50000+



Income level per *education level*

What is the breakdown of **instance weight** by **education** ?

Filtered by predicted_income: -50000, 50000+ (X)



Income level *per country of origin*

What are the values of **instance weight** by **country of birth self** ?

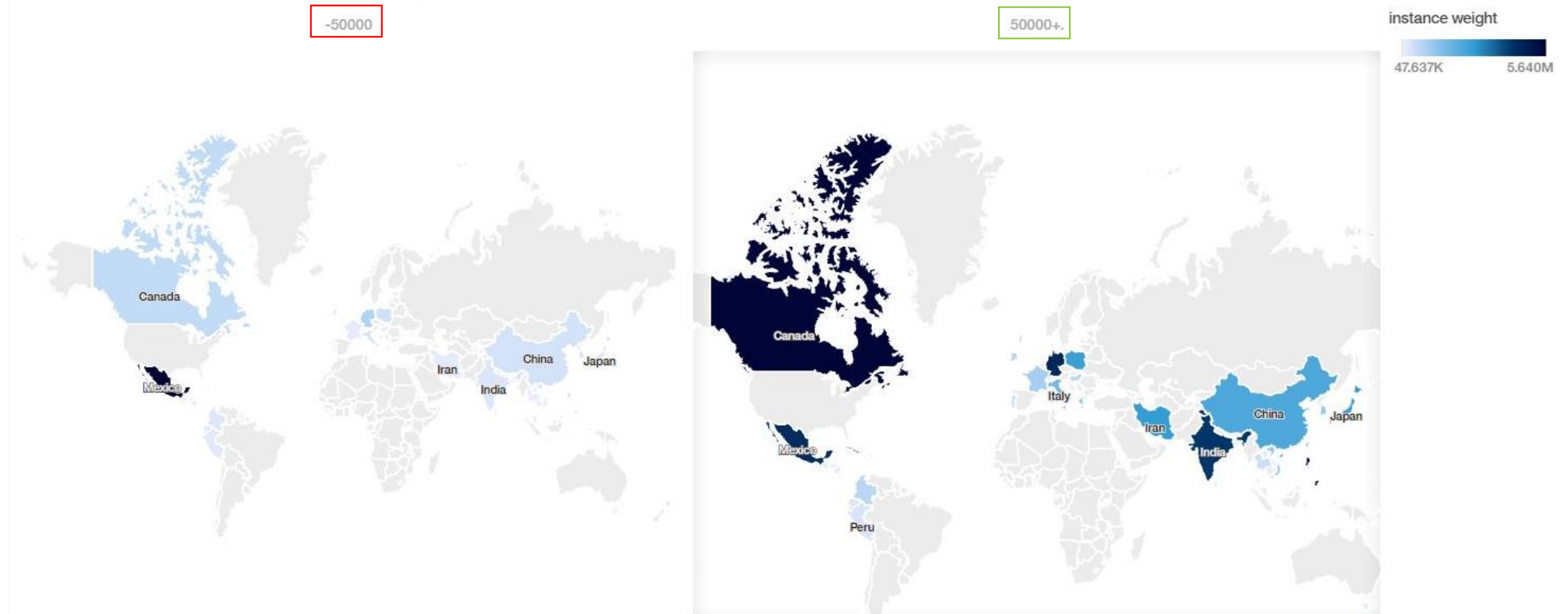
Filtered by country of birth self: 42 selected (x) and predicted_income: -50000, 50000+. (x)

-50000

50000+.

instance weight

47.637K 5.640M



Income level *per occupation*

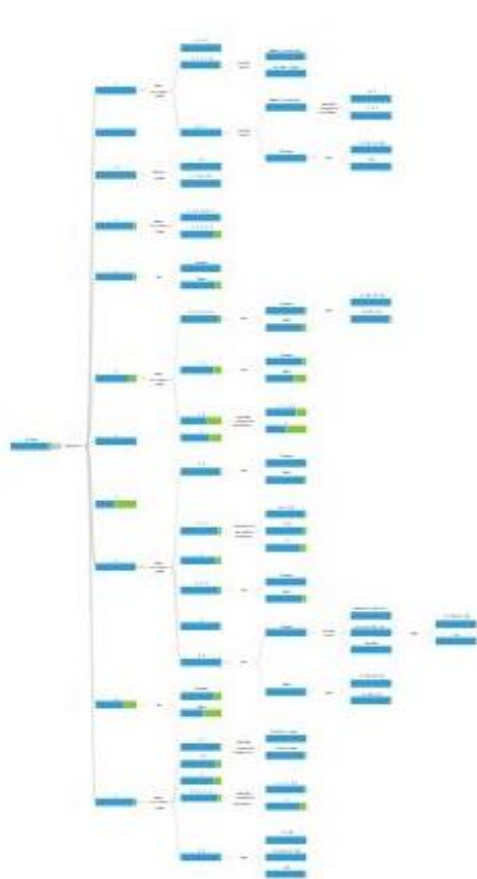
What is the breakdown of **instance weight** by **major occupation code** ?



Predict income level – *Decision tree*

What is a predictive model for `predicted_income` ?

Predictive strength: 93.7%



Predicting = -50000 ▾

Decision rules

Predicted category ▲ ▾

education = 6

age = [13, 28]; (40, 55]
tax filer status = Nonfiler
major occupation code = 6; 8; 11

tax filer status = Nonfiler
major occupation code = 6; 8; 11
education = 0

tax filer status = Nonfiler
sex = Female
major occupation code = 6; 8

age = >55
tax filer status = Nonfiler
major occupation code = 6; 8; 11

-50000 | 100%

-50000 | 100%

-50000 | 100%

-50000 | 100%

-50000 | 100%

Predicting = 50000+ ▾

Decision rules

Predicted category ▲ ▾

education = 7

detailed household and family stat =
major occupation code = 2; 9
education = 5

sex = Male
education = 10

major occupation code = 2; 9
education = 5

education = 10

50000+ | 53%

50000+ | 49%

50000+ | 44%

50000+ | 37%

50000+ | 31%

Predict income level – *Spiral*

What drives `predicted_income` ?



| Drivers | Strength | |
|---|----------|---|
| major occupation code and education | 94% | ⊕ |
| major industry code and education | 94% | ⊕ |
| weeks worked in year and education | 94% | ⊕ |
| education and industry code | 94% | ⊕ |
| education and occupation code | 94% | ⊕ |
| full or part time employment stat and education | 94% | ⊕ |
| age_group and education | 94% | ⊕ |
| education and class of worker | 94% | ⊕ |
| num persons worked for employer and education | 93% | ⊕ |
| tax filer status and education | 93% | ⊕ |
| View more | | |

Top 5 drivers

- Occupation
- Education
- Type of employment
- Age group
- Tax filler status

Dashboard

What is the summary of instance weight ?

213109131.75

instance weight (Sum)

How do the values of instance weight compare by age_cat ?



What is the breakdown of instance weight by marital status ?



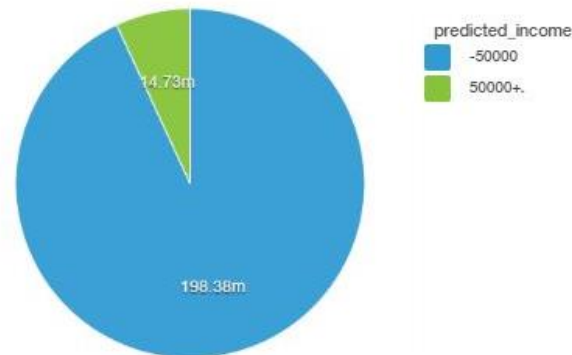
What is the breakdown of instance weight by sex ?



What are the most common values of education ?



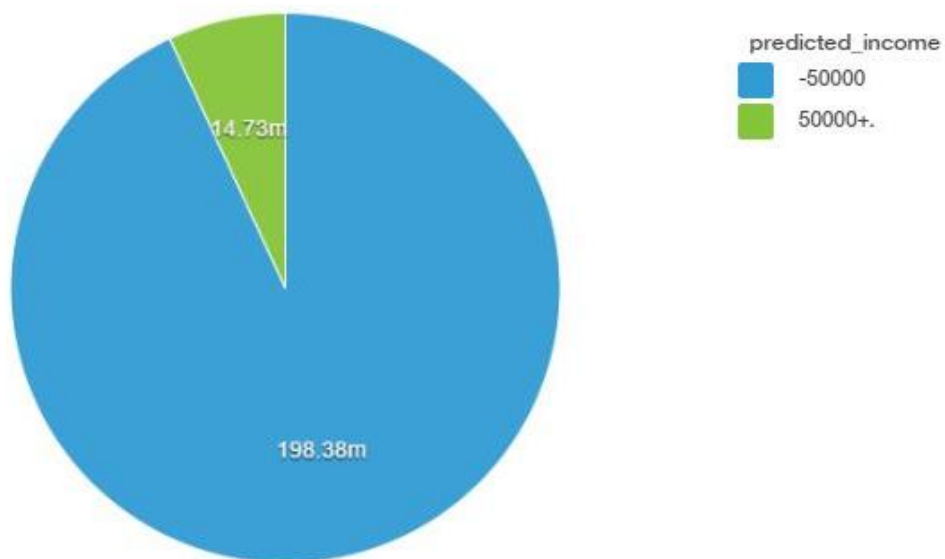
What is the breakdown of instance weight by predicted_income ?



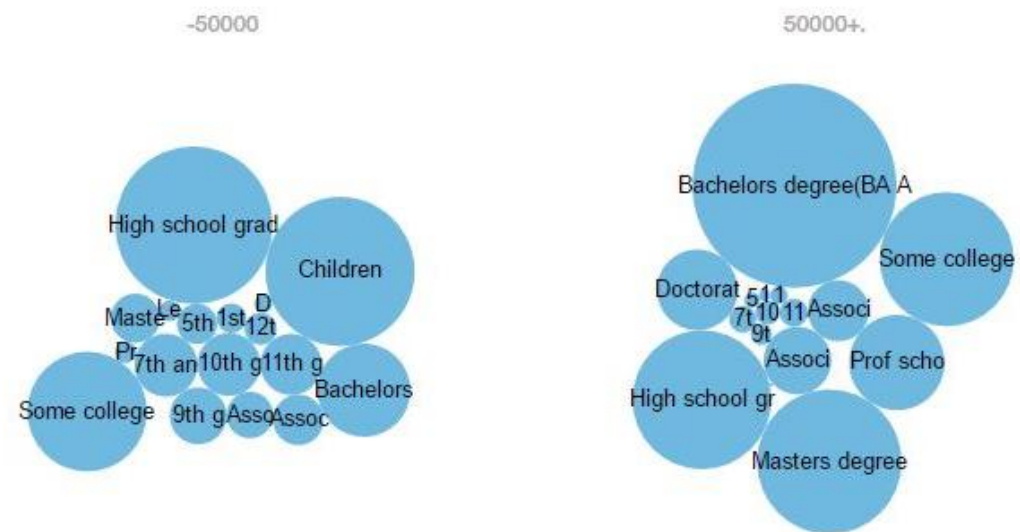
So, we won the elections...



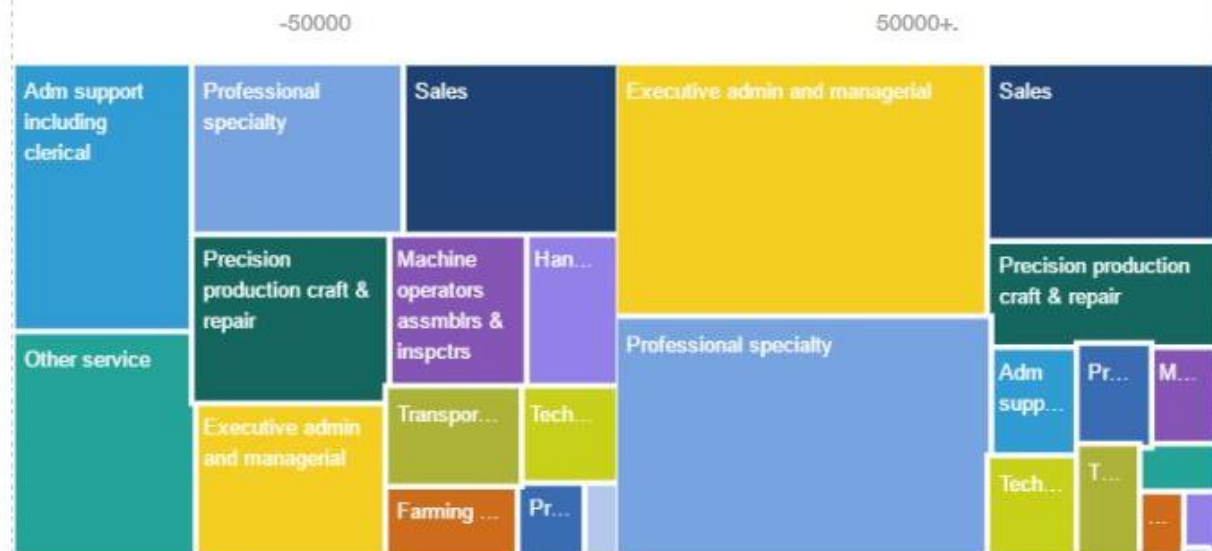
What is the breakdown of instance weight by predicted_income ?



What are the values of instance weight for each education ?



What is the breakdown of instance weight by major occupation code ?



What are the values of instance weight by country of birth self ?





...thank you!