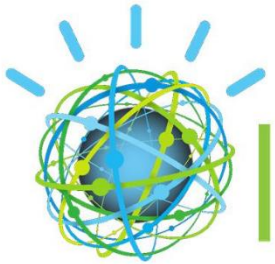
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

BUSINESS ANALYTICS
Master of Science

IBM assignment

Danai Avratoglou – BAFT1512
Katerina Kalaidopoulou – BAFT1509
Nikos Papakonstantinou – BAFT1509
Konstantinos Fraidakis – BAFT1507
Konstantinos Chronis – BAFT1502



IBM Watson

Athens, 06/06/2016

Agenda

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

Our team

Danai

Katerina

Nikos

Konstantinos F.

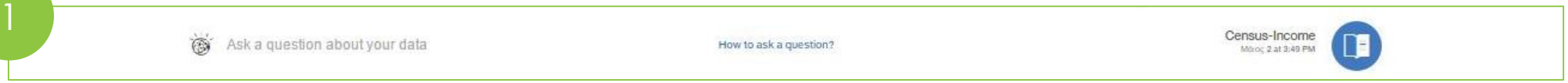
Konstantinos C.

IBM Watson

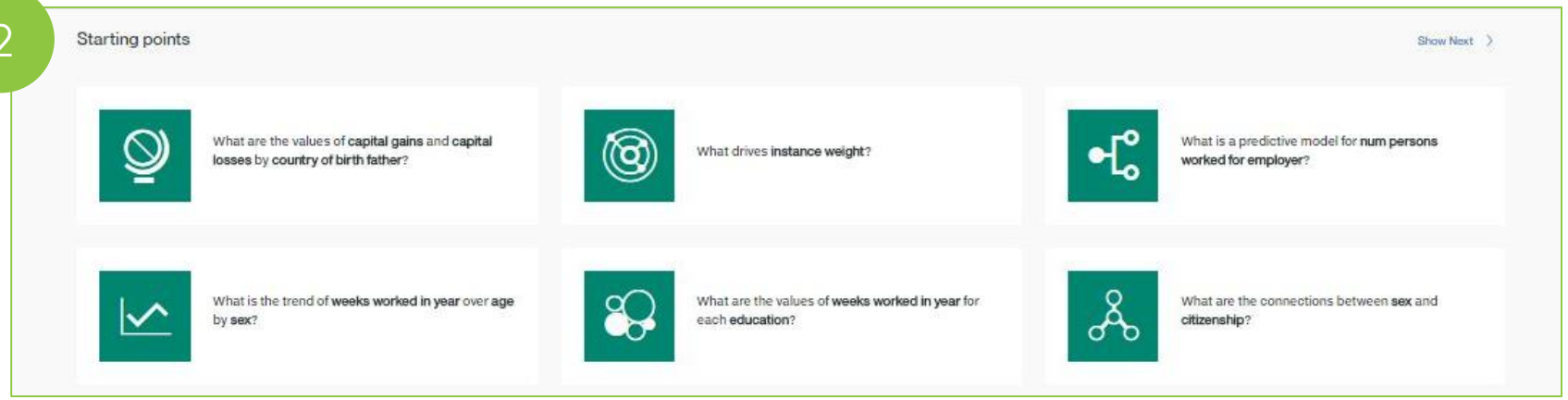
IBM

Watson Analytics contribution to data understanding

1



2



3

4

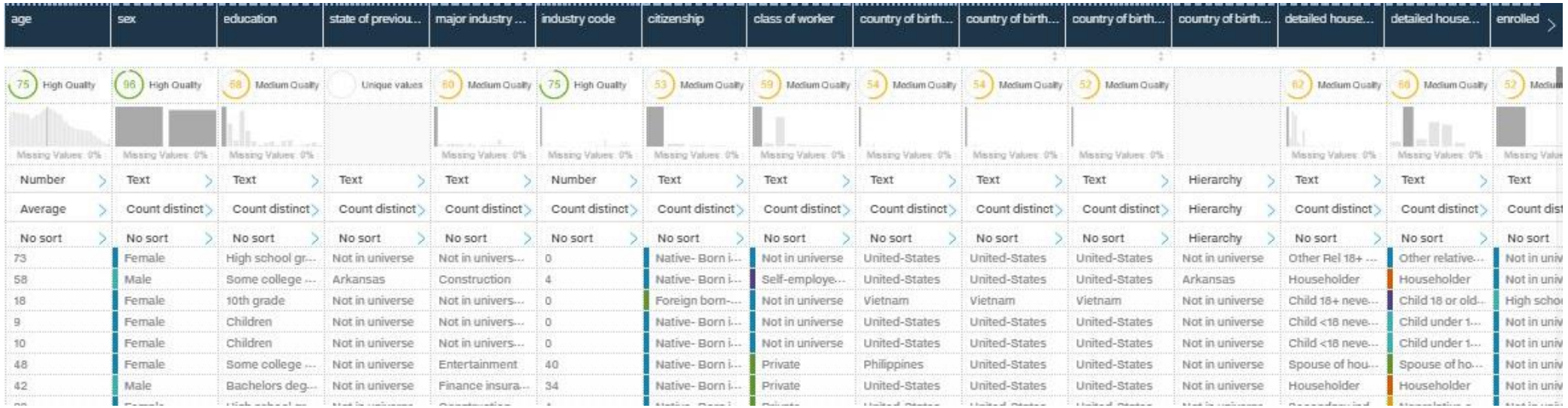
IBM Watson

IBM

Census Data*

39 Columns

146,141 Rows



age	sex	education	state of previou...	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	58 Medium Quality	52 Medium
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 dist
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 scho
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)

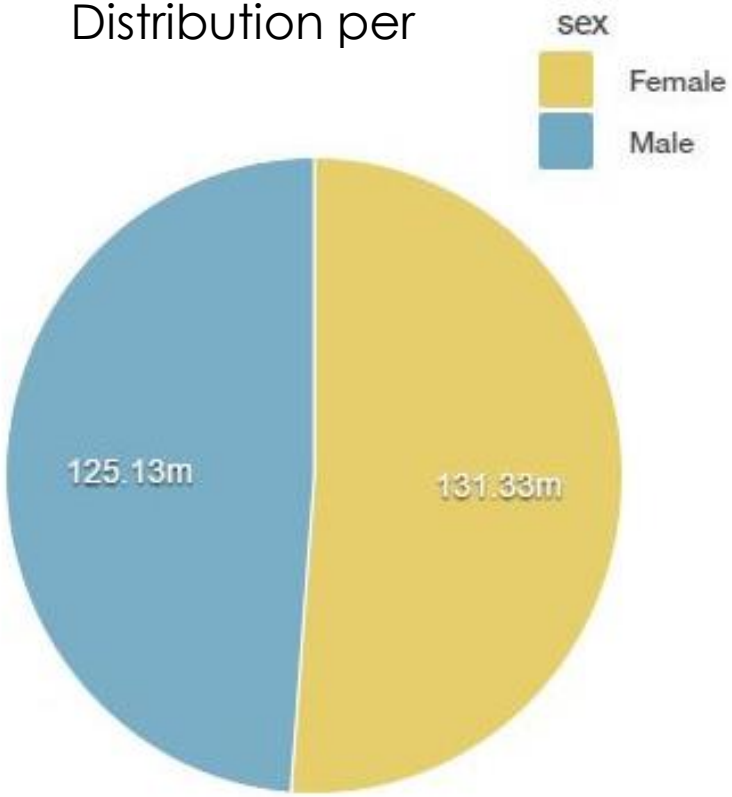
http://archive.ics.uci.edu/ml/machine-learning-databases/census-income-mld/

IBM Watson

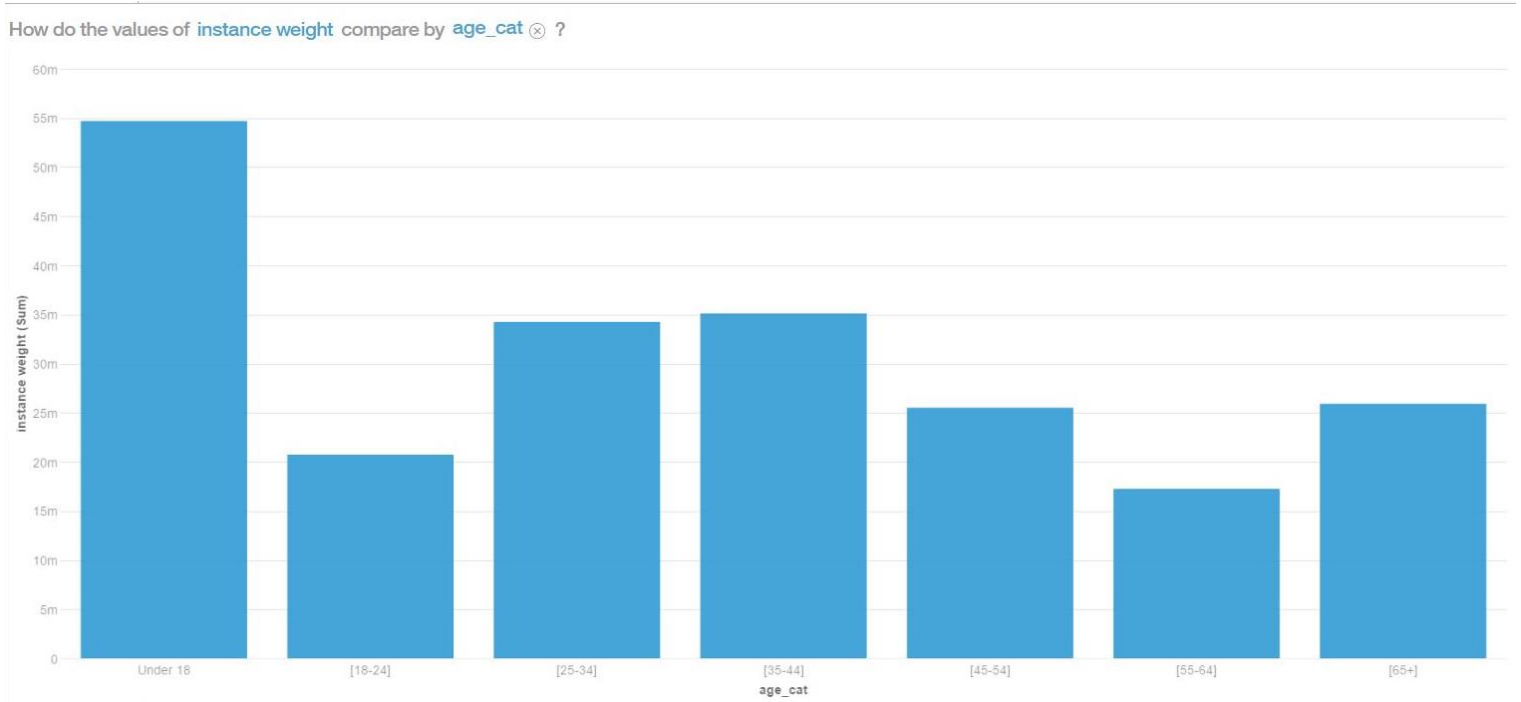
IBM

Data understanding

Distribution per



Distribution per age group



IBM Watson

IBM

Data understanding

Top Education levels

Distribution per Marital status

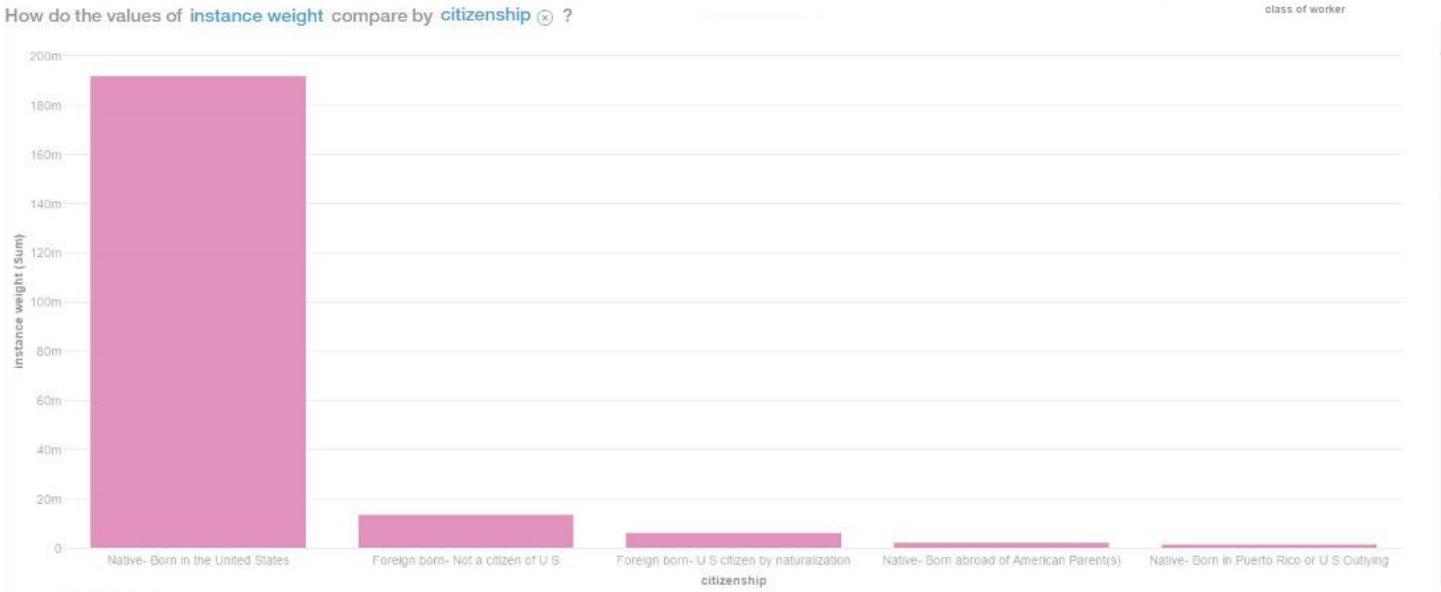
IBM Watson

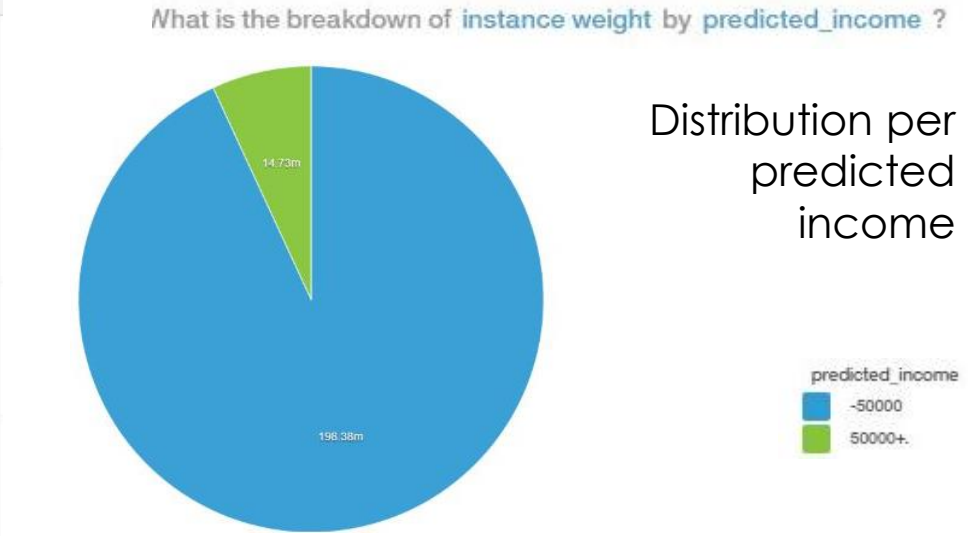
IBM

Data understanding

Distribution per class of worker

Distribution per Citizenship



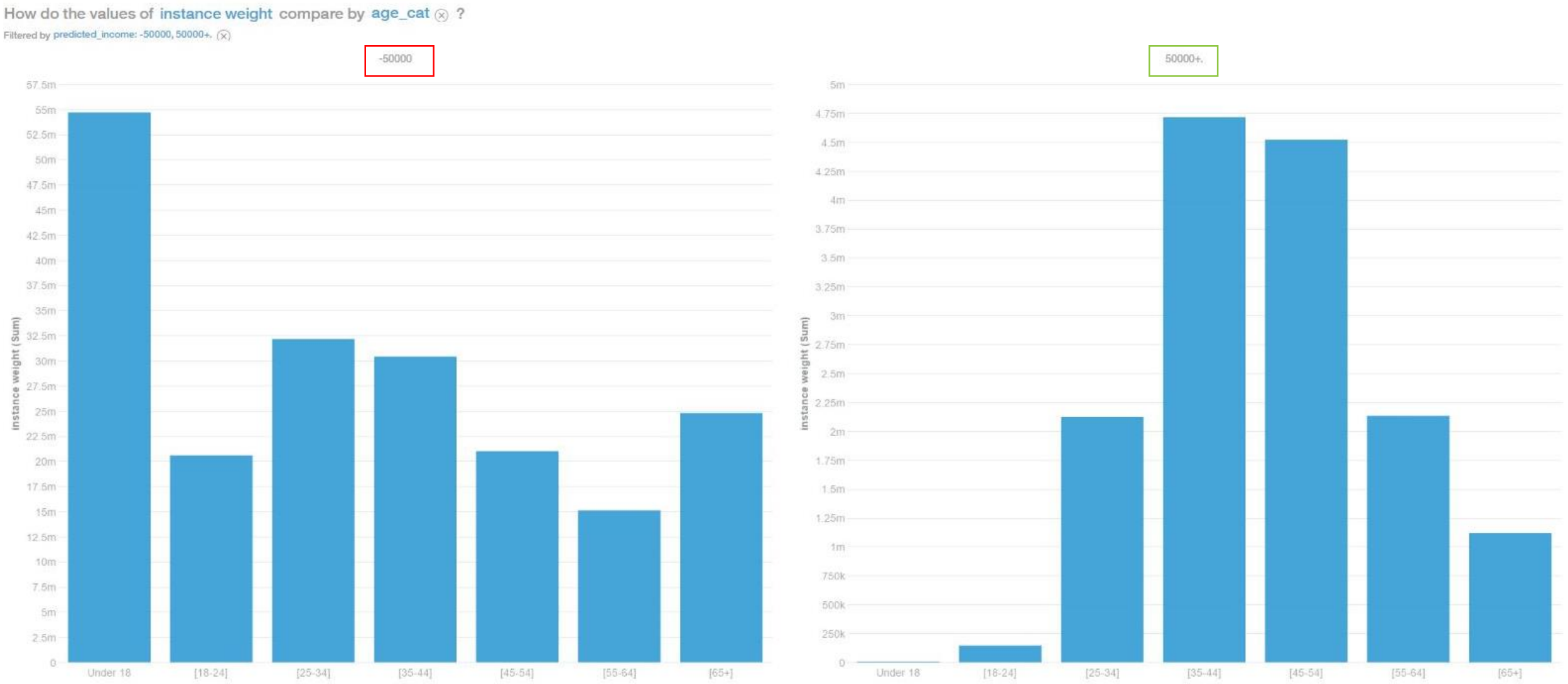


Distribution per predicted income

IBM Watson

IBM

Income level per *age group*



IBM Watson

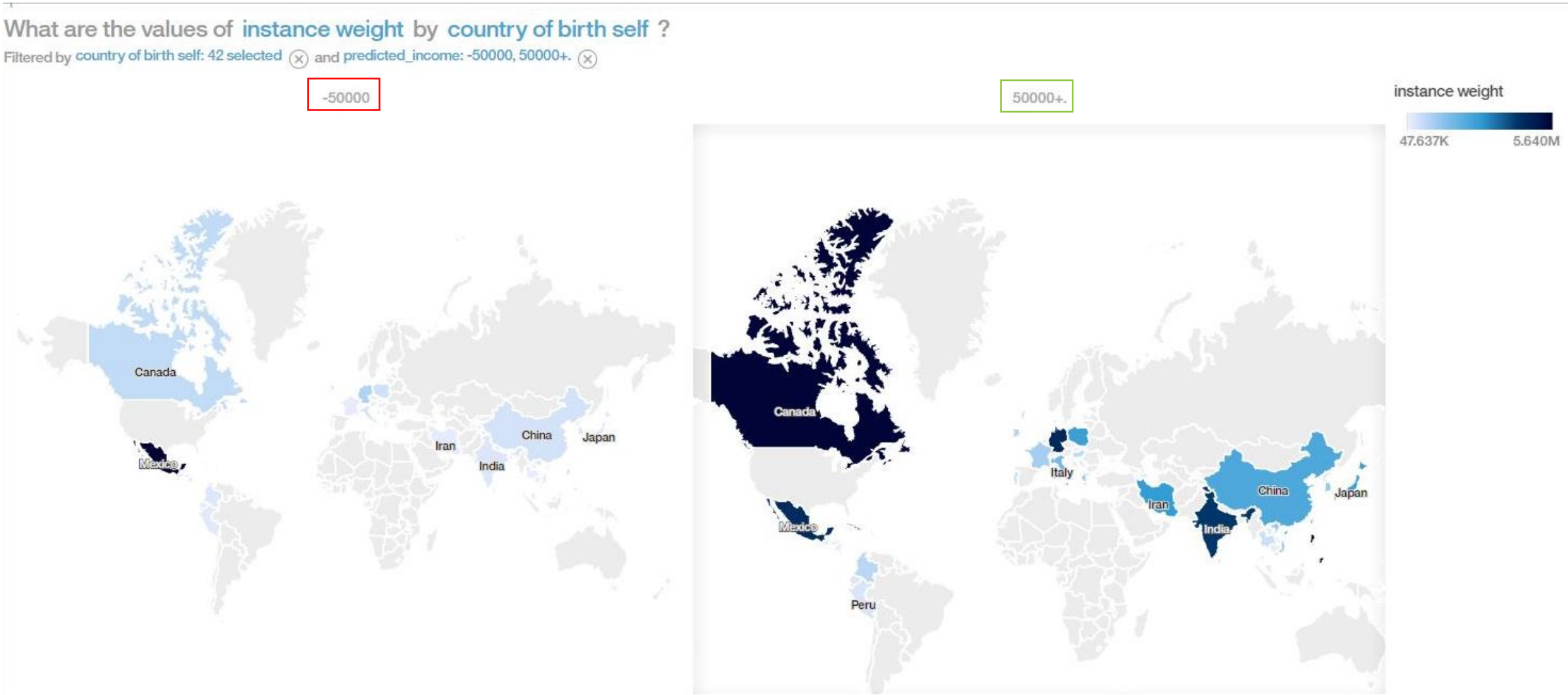
IBM

Income level per *education level*

IBM Watson

IBM

Income level *per country of origin*



IBM Watson

IBM

Income level *per occupation*

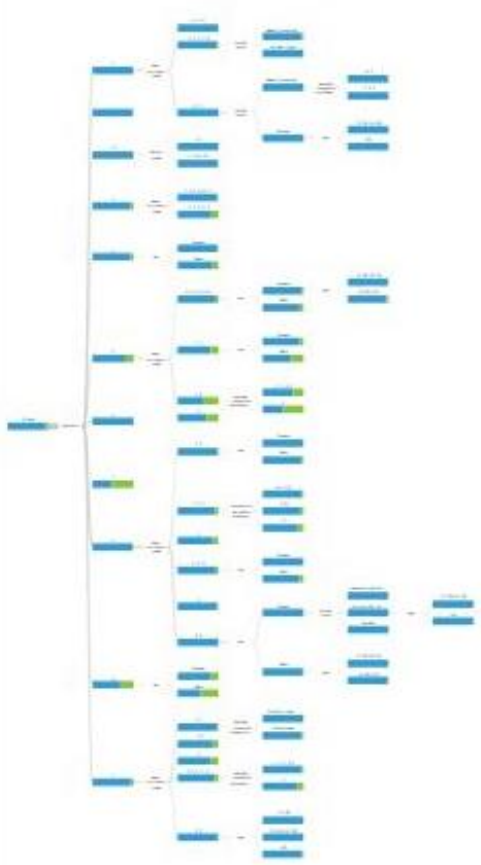
IBM Watson

IBM

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	-50000 | 100%
age = (13, 28]; (40, 55] tax filer status = Nonfiler major occupation code = 6; 8; 11	-50000 | 100%
tax filer status = Nonfiler major occupation code = 6; 8; 11 education = 0	-50000 | 100%
tax filer status = Nonfiler sex = Female major occupation code = 6; 8	-50000 | 100%
age = >55 tax filer status = Nonfiler major occupation code = 6; 8; 11	-50000 | 100%

Predicting = 50000+.

Decision rules	Predicted category
education = 7	50000+. | 53%
detailed household and family stat = major occupation code = 2; 9 education = 5	50000+. | 49%
sex = Male education = 10	50000+. | 44%
major occupation code = 2; 9 education = 5	50000+. | 37%
education = 10	50000+. | 31%

IBM Watson

IBM

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

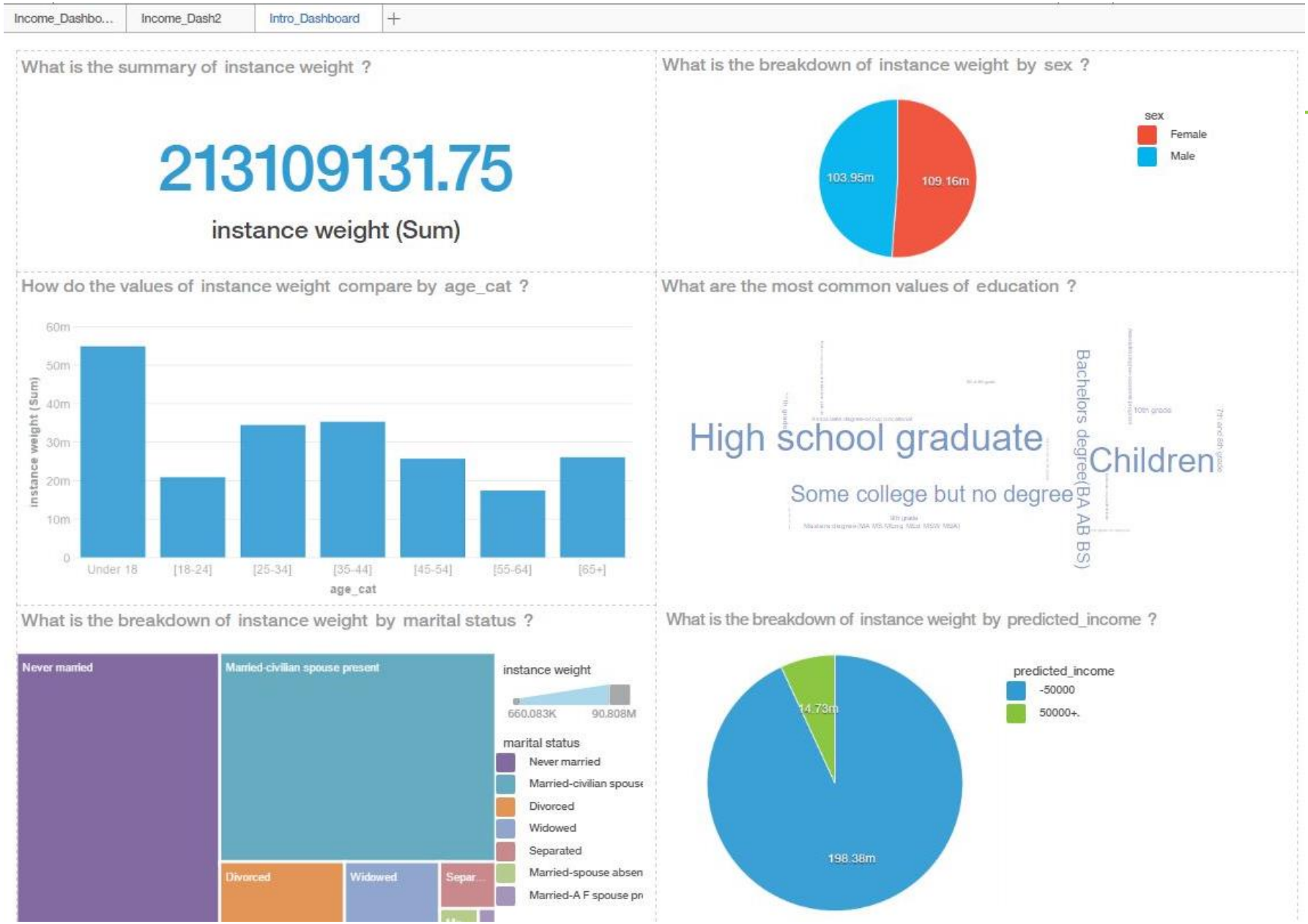
Top 5 drivers

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

IBM Watson

IBM

Dashboard

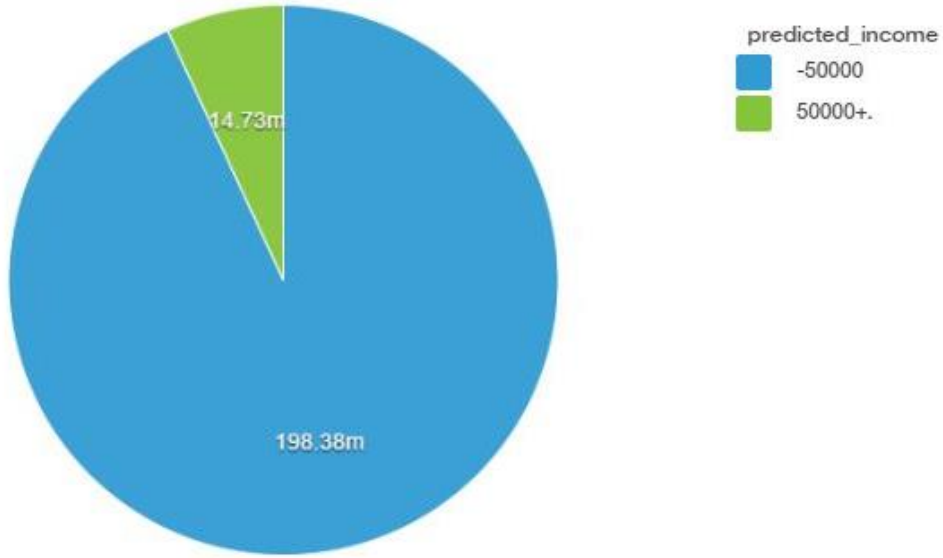


IBM Watson

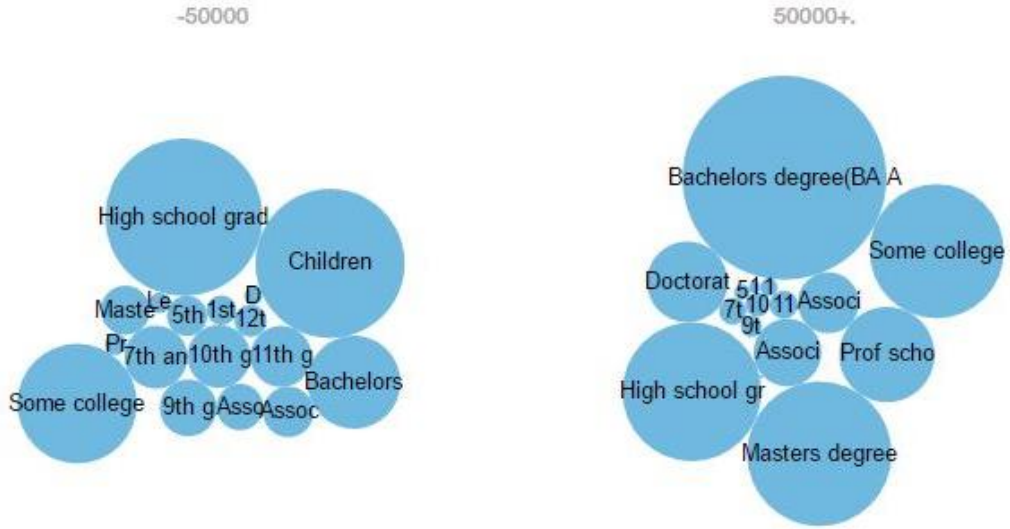
IBM

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!