Background

A multinational Academic background, grew up in Morocco, moving to France for high school. A scientific baccalaureate option mathematics, a bachelor in mathematics and currently an engineer student from IMT Atlantique.
Specializations : machine learning, deep learning, data science, cybersecurity, statistics, probability and applied mathematics.

Motivations

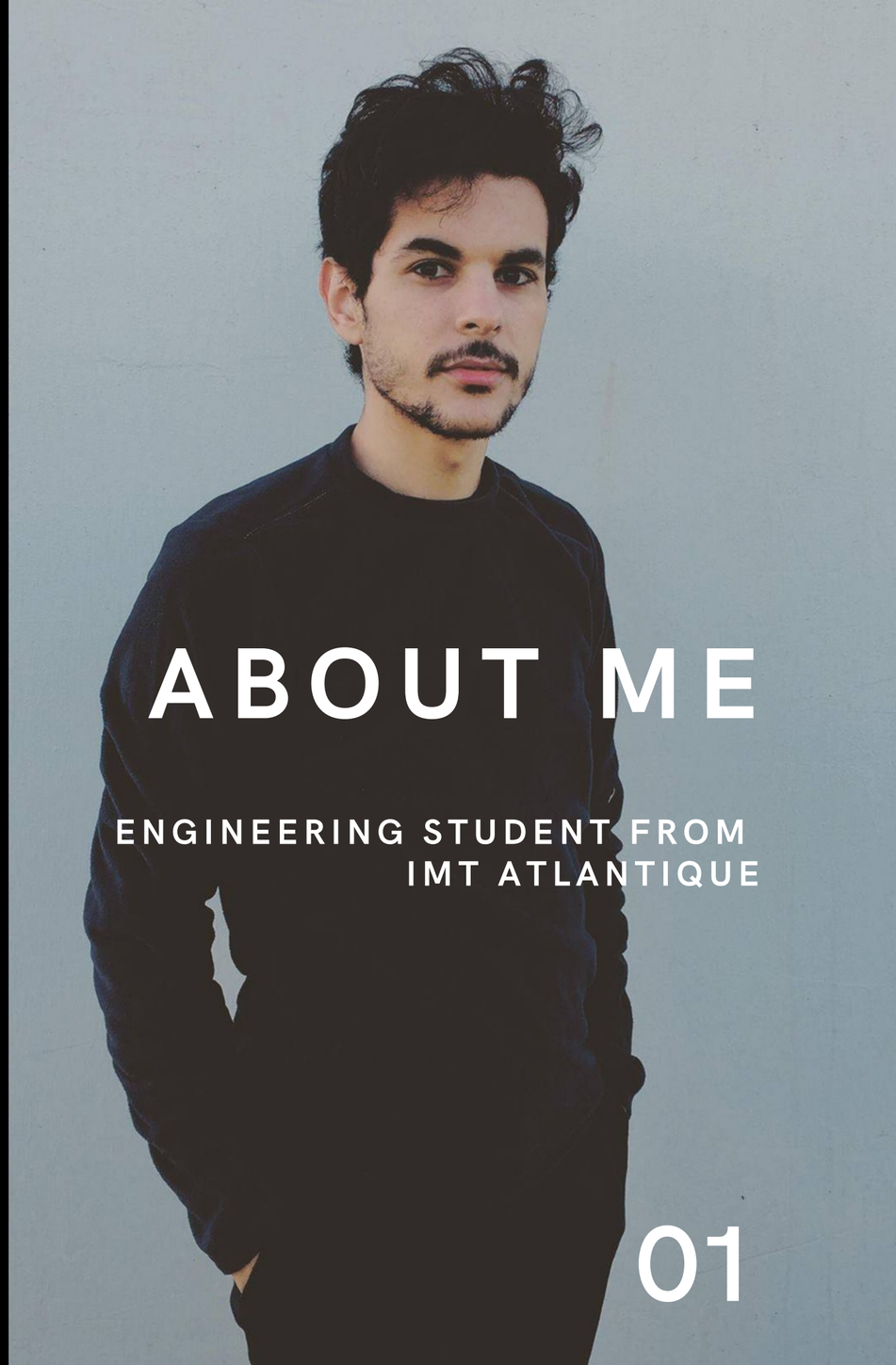
Outgoing and hard-working student, I am a quick learner curious about new technologies, open to face new challenges to develop my perspectives, and always looking for new projects to expand my skill set.

Strengths

Hardworking, Persevering, Communication skills, Writing skills, Analytical skills, Honesty, Patience, Initiative and Self-motivation.

Weaknesses

Public speaking and Self-critical/sensitive



ABOUT ME

ENGINEERING STUDENT FROM IMT ATLANTIQUE

01

MY PROUDEST ACHIEVEMENT

DIAGNOSTIC FLUX

Societe Generale's ambitions with such a project are :

- Position Sales in an advisory role
- Provide added value to customers
- Detect opportunities for rationalization, optimization and also commercial opportunities in order to improve the customer relationship.
- The other important challenge is to use data in the service of customer knowledge and therefore gain insight into customer activities and behaviors.

Context

Offering a "*diagnosis of its Cash Management flows*" to a client means offer them optimization and development opportunities, with solutions adapted to their needs. Societe Generale's ambition with such a project is to be a low-balance sheet relational banker and a service company whose objective is to complement the role of salespeople with an advisory dimension, based on an analysis of activity data and behavior of their customers

Mission

To allow sales to have a complete vision of customer behavior in terms of flow, and thus be more efficient and relevant during meetings with customers, it is necessary to have a certain number of indicators and graphs that help to provide a first automated solution to this business need. As a result of data mining, several widgets are produced.

Results

Offer a finalized Dashboard which brings together the various widgets and the most importantly the one that responds to sales requests.
This Dashboard must be readable and easy to use

03 PROBLEM STATEMENT

IDENTIFYING CHARACTERISTICS THAT ARE ASSOCIATED WITH A PERSON MAKING MORE OR LESS THAT 50 000$

04 DIAGNOSIS OF THE CURRENT SITUATION

DISTRIBUTION OF ENTRIES ACCORDING TO OUR TWO LABELS

Distribution of entries according to our two labels

<= 50 000 $	> 50 000 $
93.79 %	6.21 %

The original dataset contains a distribution of 6.21% entries labeled with >50k and 93.79% entries labeled with <=50k.

The Census Income dataset has 199 523 entries.

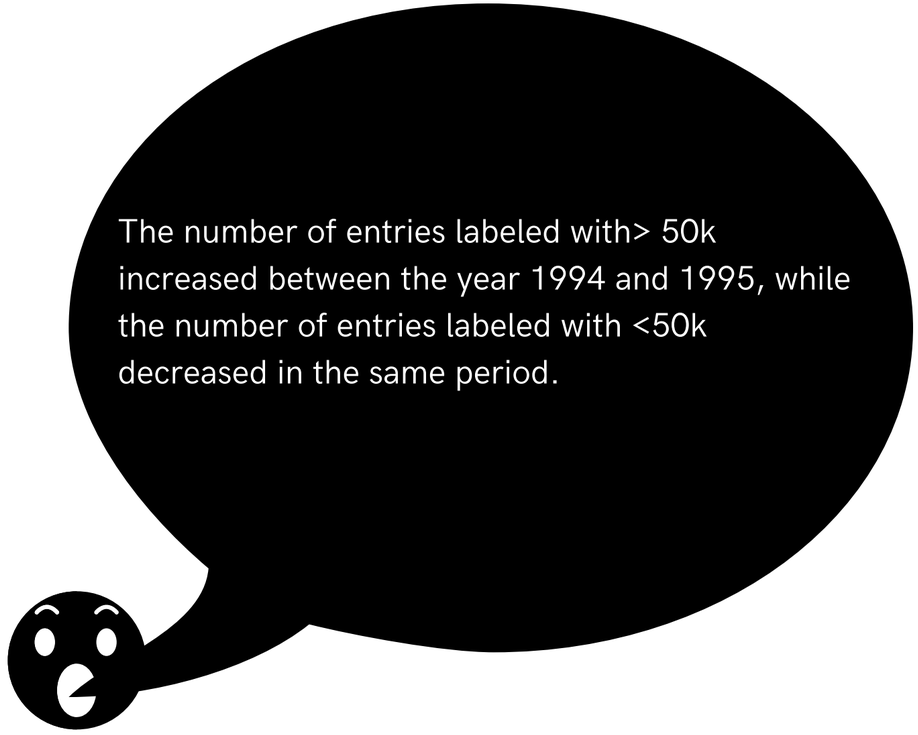
05 DIAGNOSIS OF THE CURRENT SITUATION

YEAR DISTRIBUTION AMONG THE ENTRIES IN OUR DATASET

Distribution of entries according to our two labels per year

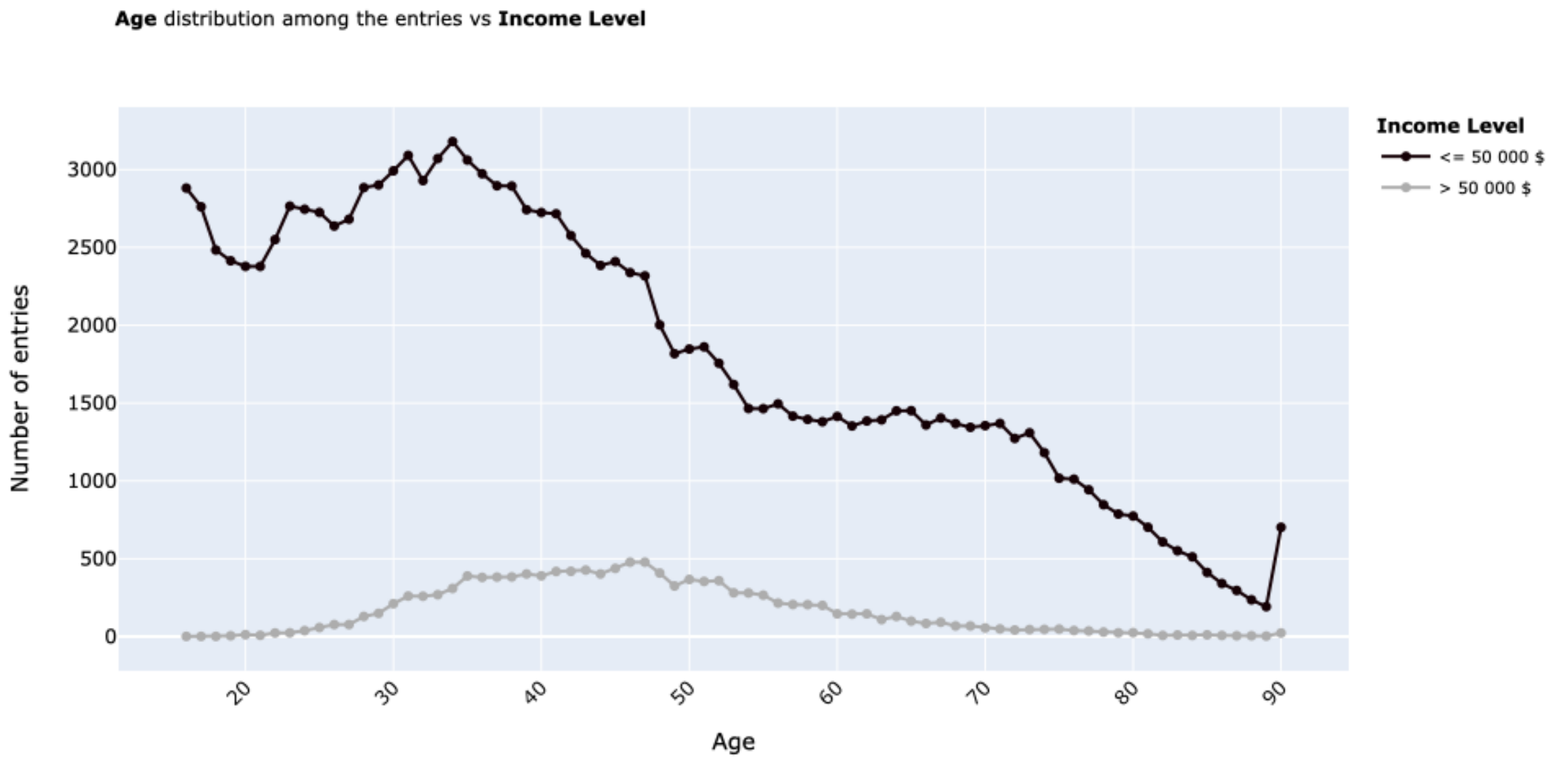
year	<= 50 000 $	> 50 000 $
94	93988	5839
95	93153	6543

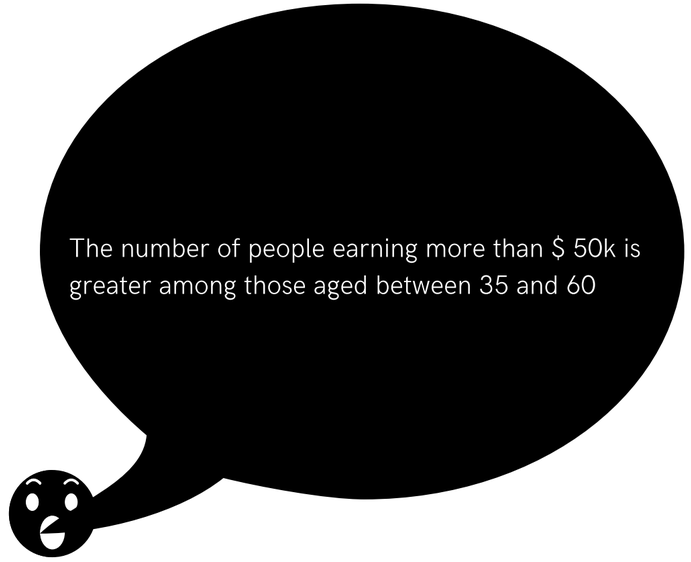
The number of entries labeled with> 50k increased between the year 1994 and 1995, while the number of entries labeled with <50k decreased in the same period.



06 DIAGNOSIS OF THE CURRENT SITUATION

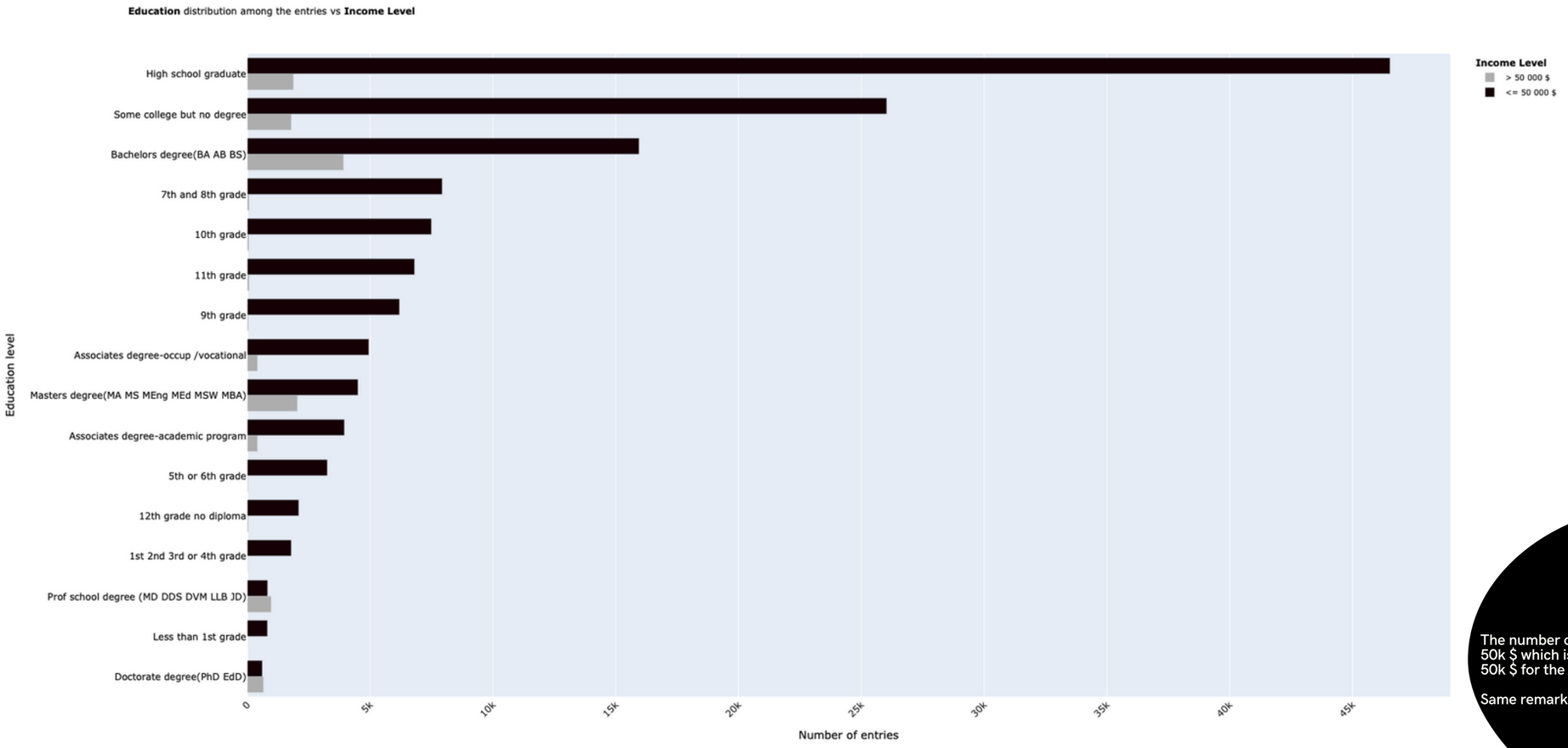
AGE DISTRIBUTION AMONG THE ENTRIES IN OUR DATASET





07 DIAGNOSIS OF THE CURRENT SITUATION

EDUCATION DISTRIBUTION AMONG THE ENTRIES IN OUR DATASET



The number of people having a Phd degree earn more than 50k $ which is slightly larger than those who earn less than 50k $ for the same level of education.

Same remark for people having prof school degree.

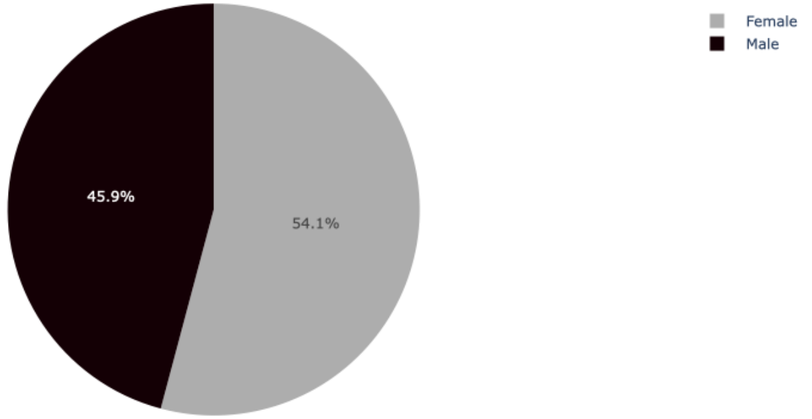
08 DIAGNOSIS OF THE CURRENT SITUATION

SEX DISTRIBUTION AMONG THE ENTRIES IN OUR DATASET

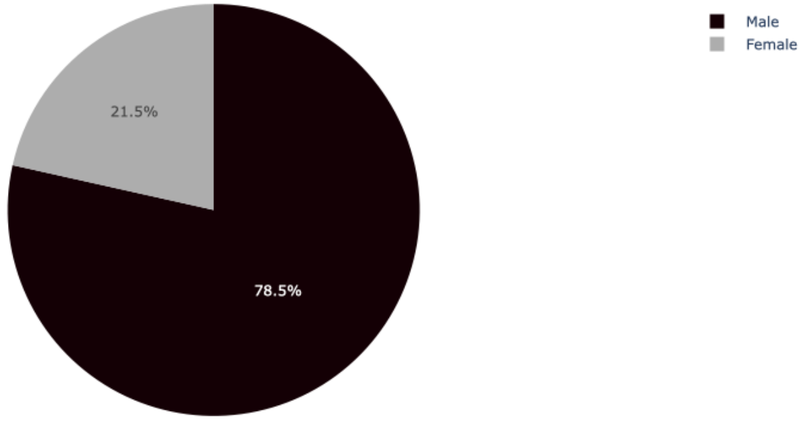
Sex distribution among the entries vs **Income Level**

sex	<= 50 000 $	> 50 000 $
Female	101321	2663
Male	85820	9719

Sex distribution among the entries vs **Income Level: <= 50 000 $**



Sex distribution among the entries vs **Income Level: > 50 000 $**

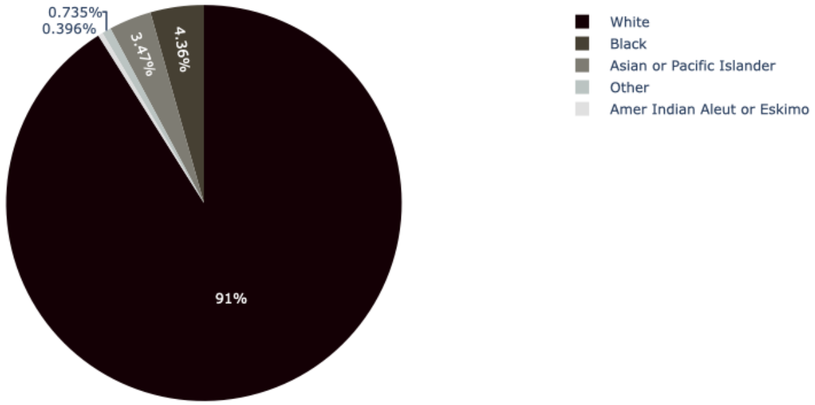


- The proportion of men earning a salary over 50k is significantly higher than that of women. It reached a level of 78.5% between the years 1994 and 1995 against only 21.5% for women.
- For a salary lower than 50k the proportions are closer although the proportion of men is slighty greater than that of women.

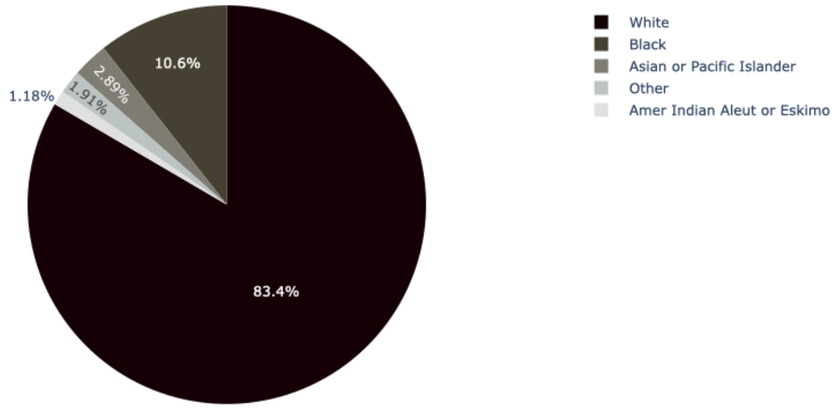
09 DIAGNOSIS OF THE CURRENT SITUATION

RACE DISTRIBUTION AMONG THE ENTRIES IN OUR DATASET

Race distribution among the entries vs **Income Level: > 50 000 $**

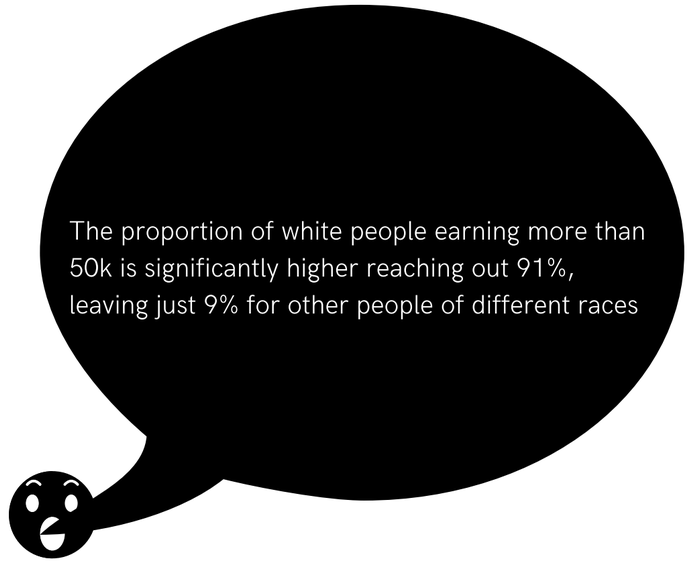


Race distribution among the entries vs **Income Level: <= 50 000$**



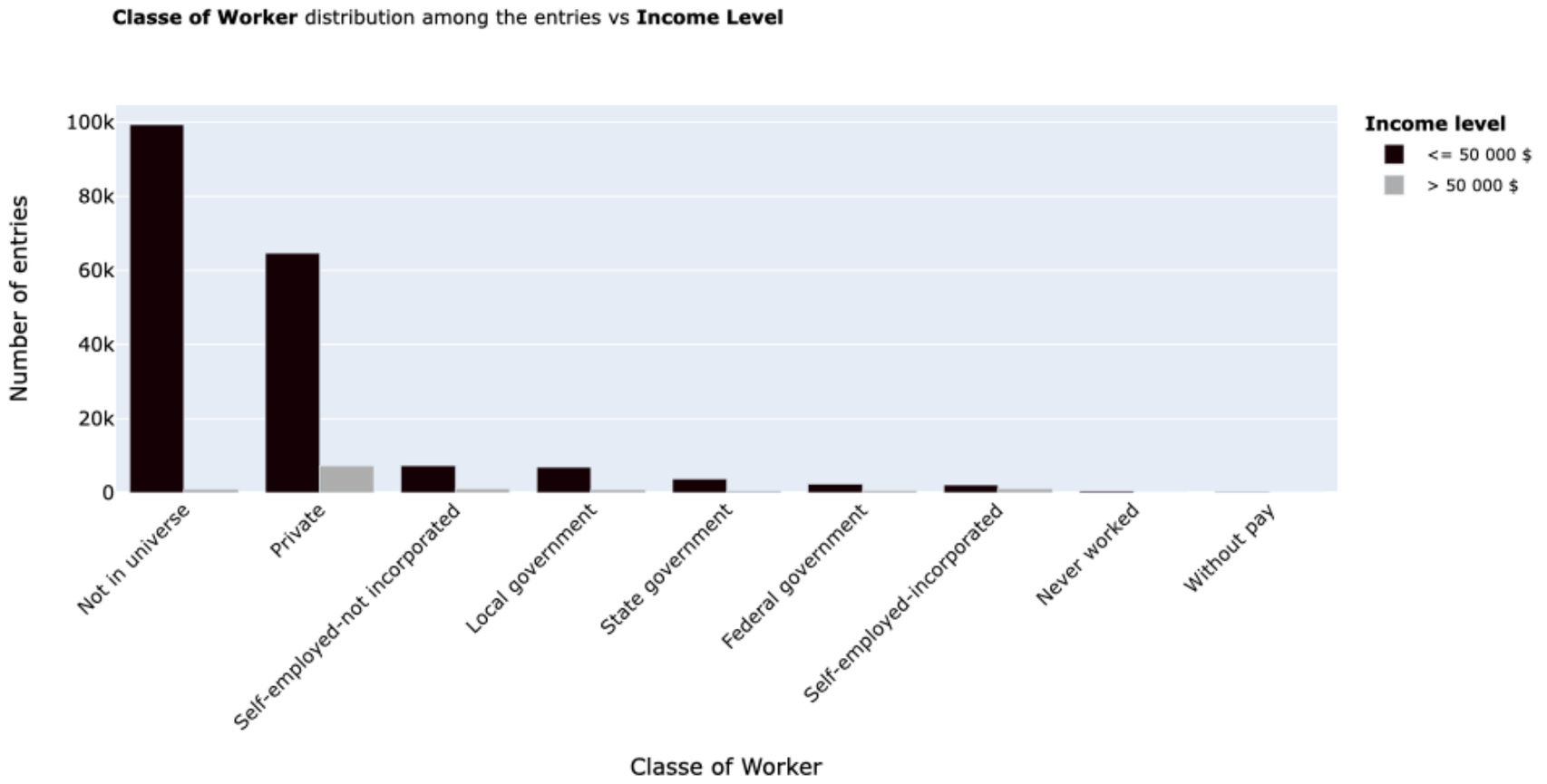
Race distribution among the entries vs **Income Level**

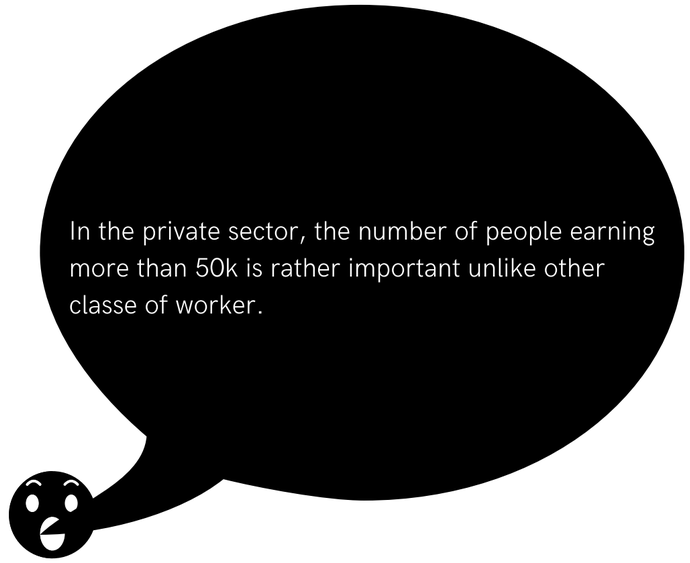
race	<= 50 000 $	> 50 000 $
White	156093	11272
Black	19875	540
Asian or Pacific Islander	5405	430
Other	3566	91
Amer Indian Aleut or Eskimo	2202	49



DIAGNOSIS OF THE CURRENT SITUATION

CLASSE OF WORKER DISTRIBUTION AMONG THE ENTRIES

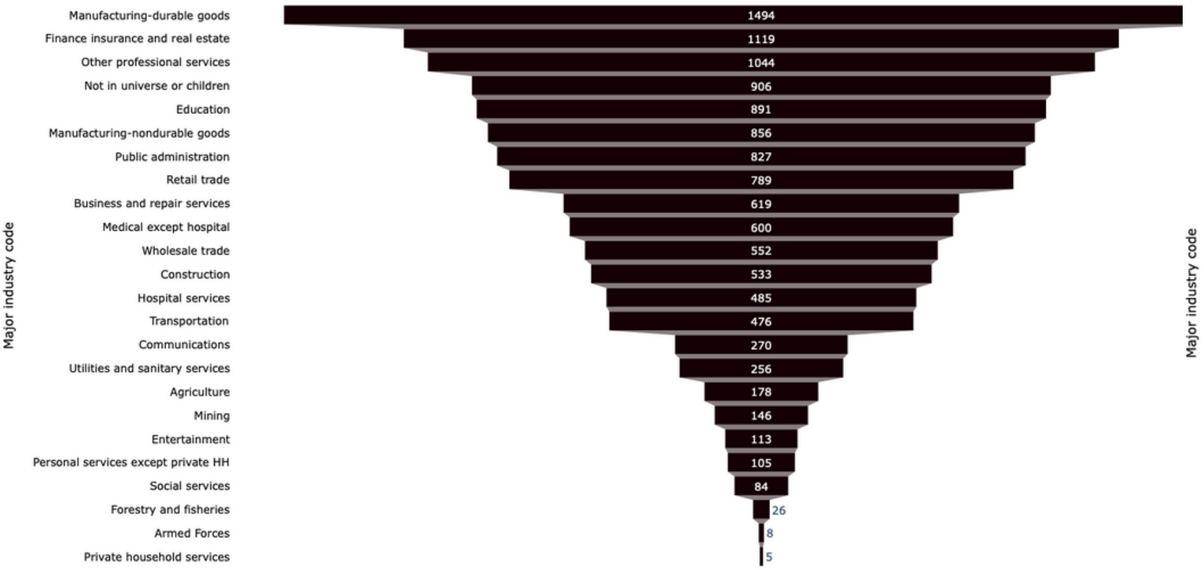




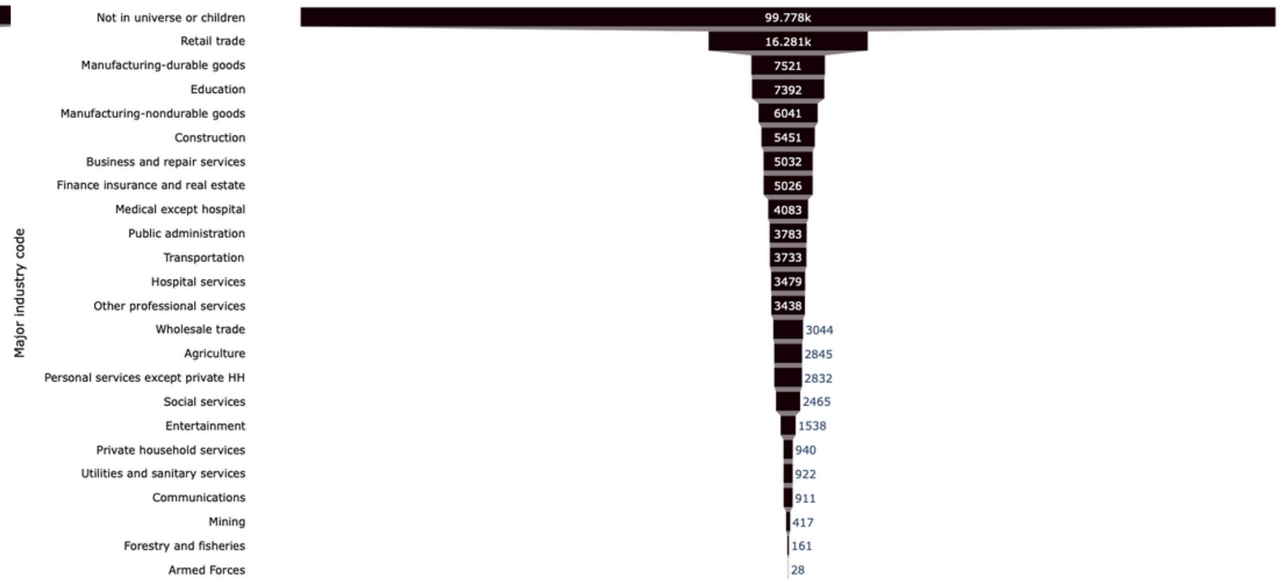
11 DIAGNOSIS OF THE CURRENT SITUATION

MAJOR INDUSTRY CODE DISTRIBUTION AMONG THE ENTRIES

Major industry code distribution among the entries vs **Income Level: > 50 000$**

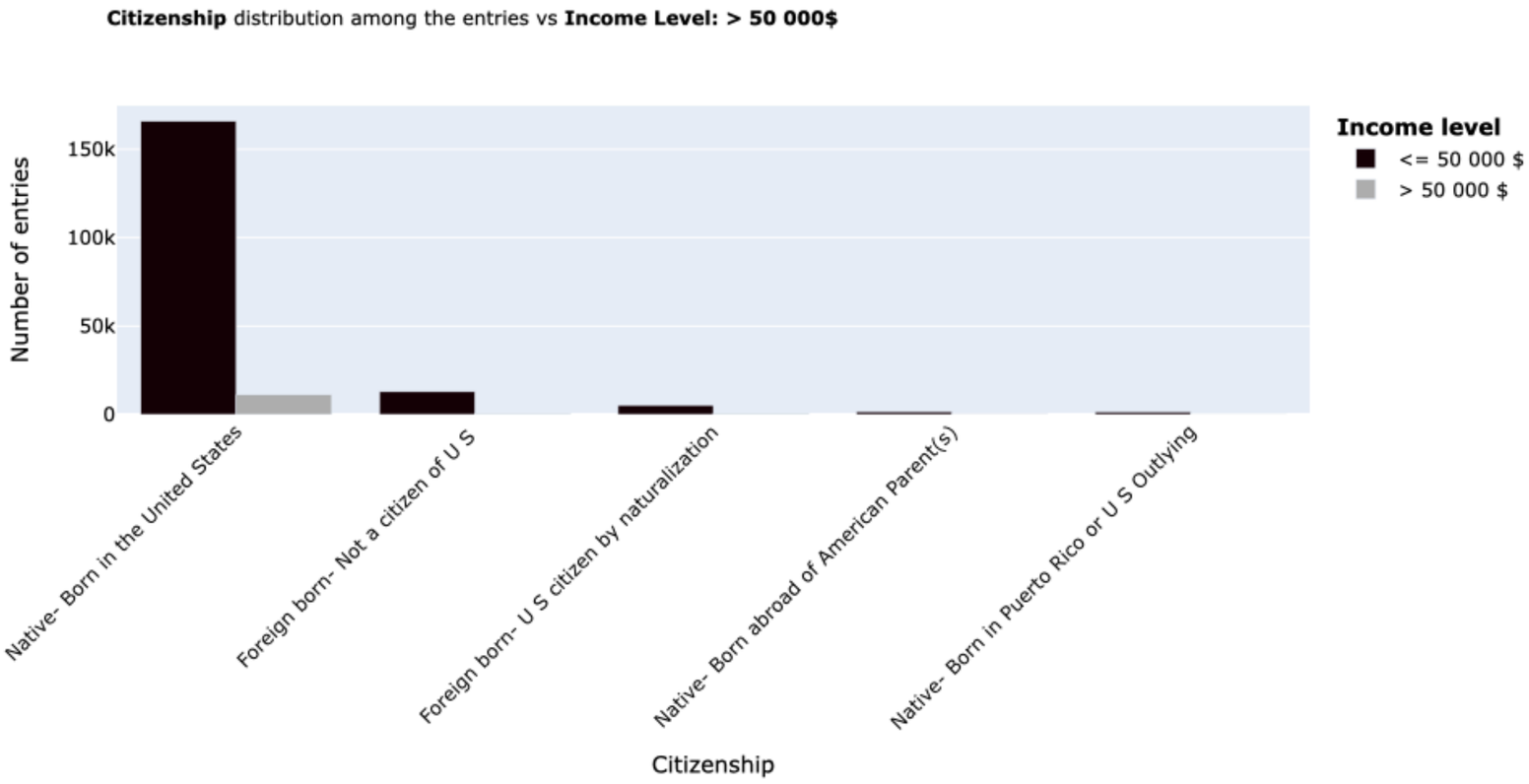


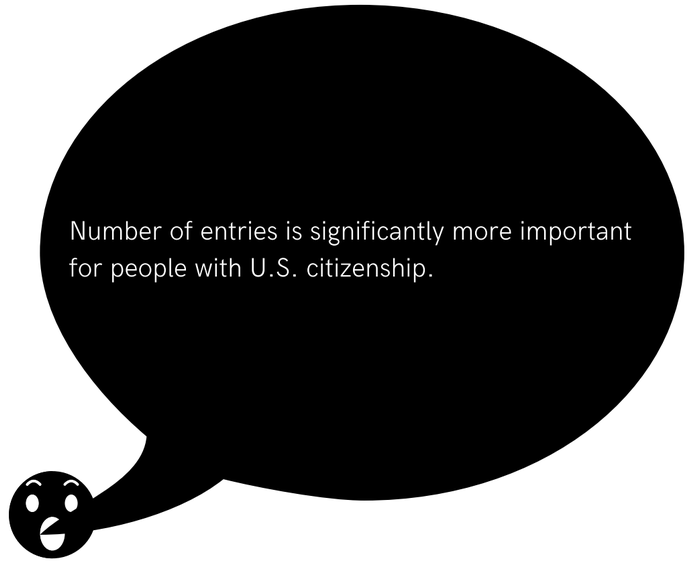
Major industry code distribution among the entries vs **Income Level: <= 50 000$**



DIAGNOSIS OF THE CURRENT SITUATION

CITIZENSHIP DISTRIBUTION AMONG THE ENTRIES IN OUR DATASET





Models

GaussianNB

Model: **Gaussian NB**

Metric	Features	Features Selection
Accuracy	0.79	0.87
Precision	0.2	0.26

Decision Tree Classifier

Model: **Decision Tree Classifier**

Metric	Features	Features Selection
Accuracy	0.93	0.93
Precision	0.45	0.44

Logistic Regression

Model: **Logistic Regression**

Metric	Features	Features Selection
Accuracy	0.95	0.95
Precision	0.7	0.68

Neural Network