COVID-19: Occupations with the Highest Risk

George Sapp
Matthew Lopez
Shahzad Ansari
Nabeel Usmani
Ken Ho

Data Boot Camp
Aug 18 2020

Summary Slide

- Analysis of Covid risk scores and determination which occupations have the highest risk.
- When Covid pandemic started, many non-essential businesses were shuttered and the news announced that many health care workers, especially in the hot spots were at highest risk of contracting Covid. At the same time, occupations such as office workers were determined to be at lowest risk if they can work remotely.
- This study assessed if there are indeed specific occupation types that are more susceptible to contracting Covid, and if salary even matters and for the occupations that are most at risk, what proportion of the working population they make up.
- The results will help decide how to better protect the occupations most at risk, determine the optimal levels and durations of services rendered while keeping the public safe and determine the frequency of Covid testing that will be required.
- Fortunately, the data sets and analytical tools available for this analysis were able to answer many of the critical public health questions posed above.

Analytical Question and Data Sources

- Question: Are there occupations more susceptible to Covid, and do salary matter, and for those that are most at risk, what proportion of the working population do they make up?

- Data Sources Needed:
  - Occupations
  - Employment Count
  - Mean Salaries

  The US Bureau of Labor Statistics was a good source of the above data

  - Covid Risk Scores

  Much more difficult to locate. Initially found data for England and Wales. Finally found the US Department of Labor's O'Net Database

Data Cleanup and Exploration

1. Start with file ***Occupation Salary Population and Type.*** Sourced from the US Bureau of Labor Statistics
2. There are three other files with physical job attributes that will need to be appended to the ***Occupation Salary Population and Type*** file to create the value Covid Risk Score
   a. ***Contact_With_Others ,*** how much does this job require the worker to be in contact with others to perform it?
   b. ***Physical_Proximity,*** to what extent does this job require the worker to perform tasks in close physical proximity to others?
   c. ***Exposure_to_Disease_and_Infection***, how often does this job require exposure to hazardous conditions?

Sourced from the US Department of Labor's O'Net Database

- Each of these three files has a field of ____ Context, which is the risk score specific to that physical job attribute. The risk score is between 0 and 100, with 100 representing the highest possible risk
- Join these three files together into one by the field Code, which is the occupation code specified by the Department of Labor.
- The resulting file should be named ***Overall Risk Score***. Create a new field Overall Risk Score, which is calculated as the average value of each of the three ____ Context.

Data Cleanup and Exploration … cont'd.

1. Match the ***Overall Risk Score*** file to the ***Occupation Salary Population and Type*** file using the Occupation field to create a new file named ***Occupation Risk Score***. Some of the Occupation may not match and all such records will be dropped.
2. Create a new field, PCT_TOT_EMP, which is the % of total employees for each occupation
3. This ***Occupation Risk Score*** file will now have six fields as such:
    a. OCCUPATION
    b. TOT_EMP for total count of employees in this occupation
    c. Health/Non Health, which is a designation of whether the occupation is Health versus Non Health
    d. MEAN_SALARY
    e. Overall Risk Score
    f. PCT_TOT_EMP

Challenges with the data

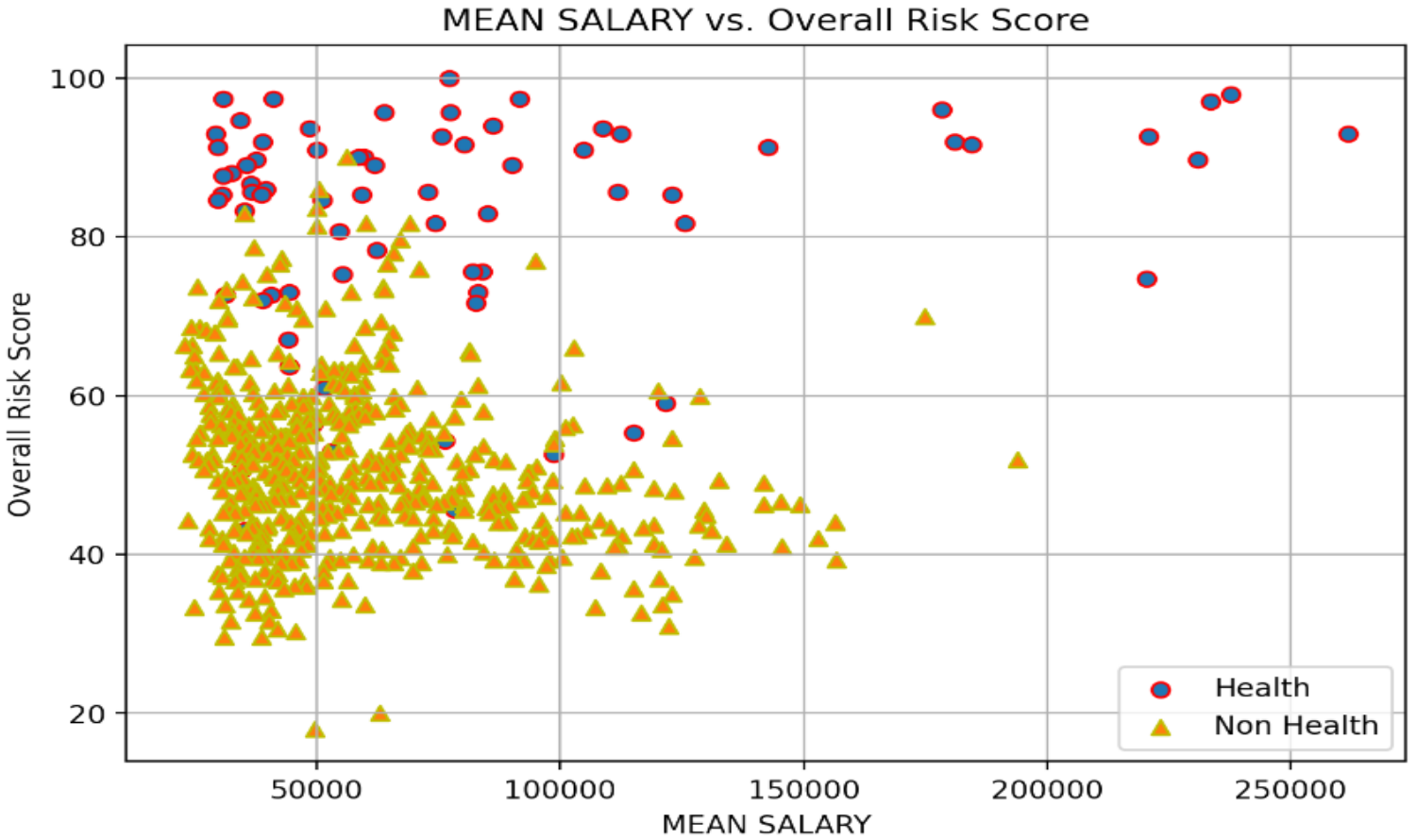
- Occupation data matching by name is not as convenient but fortunately, the data files have the same parent source, the US Department of Labor.
- Plotting the OCCUPATION name (some are long) by Overall Risk Score was challenging due to the space limitation and the 600+ occupation types, so we cut off 300 occupations with a low employment count
- Subject does not have a widely available dataset as one would expect from this critical topic

Data Analysis

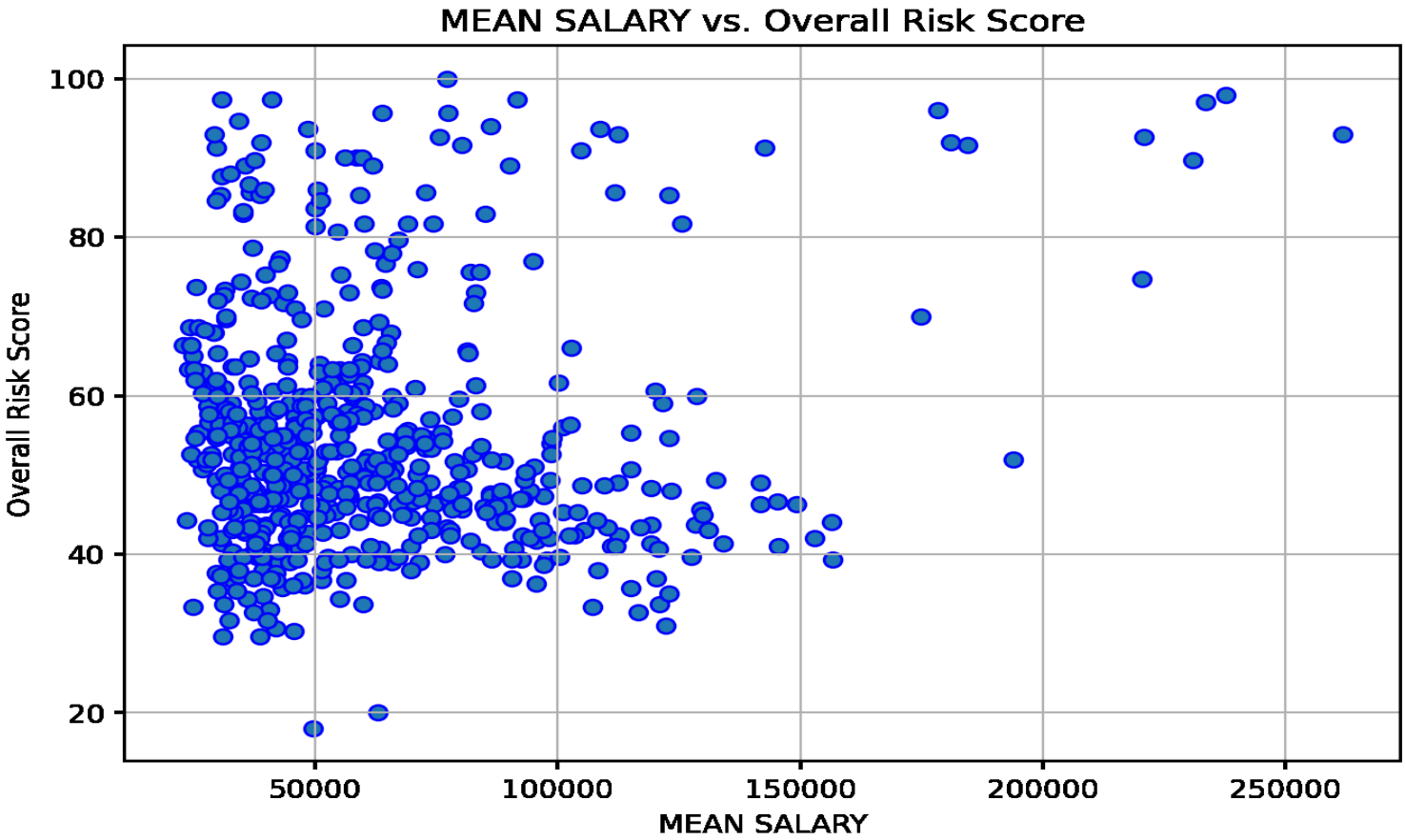
1. Plot the MEAN_SALARY to the Overall Risk Score
2. Plot the TOT_EMP to the Overall Risk Score
3. Plot the PCT_TOT_EMP to the Overall Risk Score
4. Calculate the median of the PCT_TOT_EMP
5. Drop the records with values below this median
6. Then, for the remaining records, plot the OCCUPATION to the Overall Risk Score. Color Code the OCCUPATION name by Health (red color) versus Non Health (green color)

<u>Interesting figures from the analysis:</u>

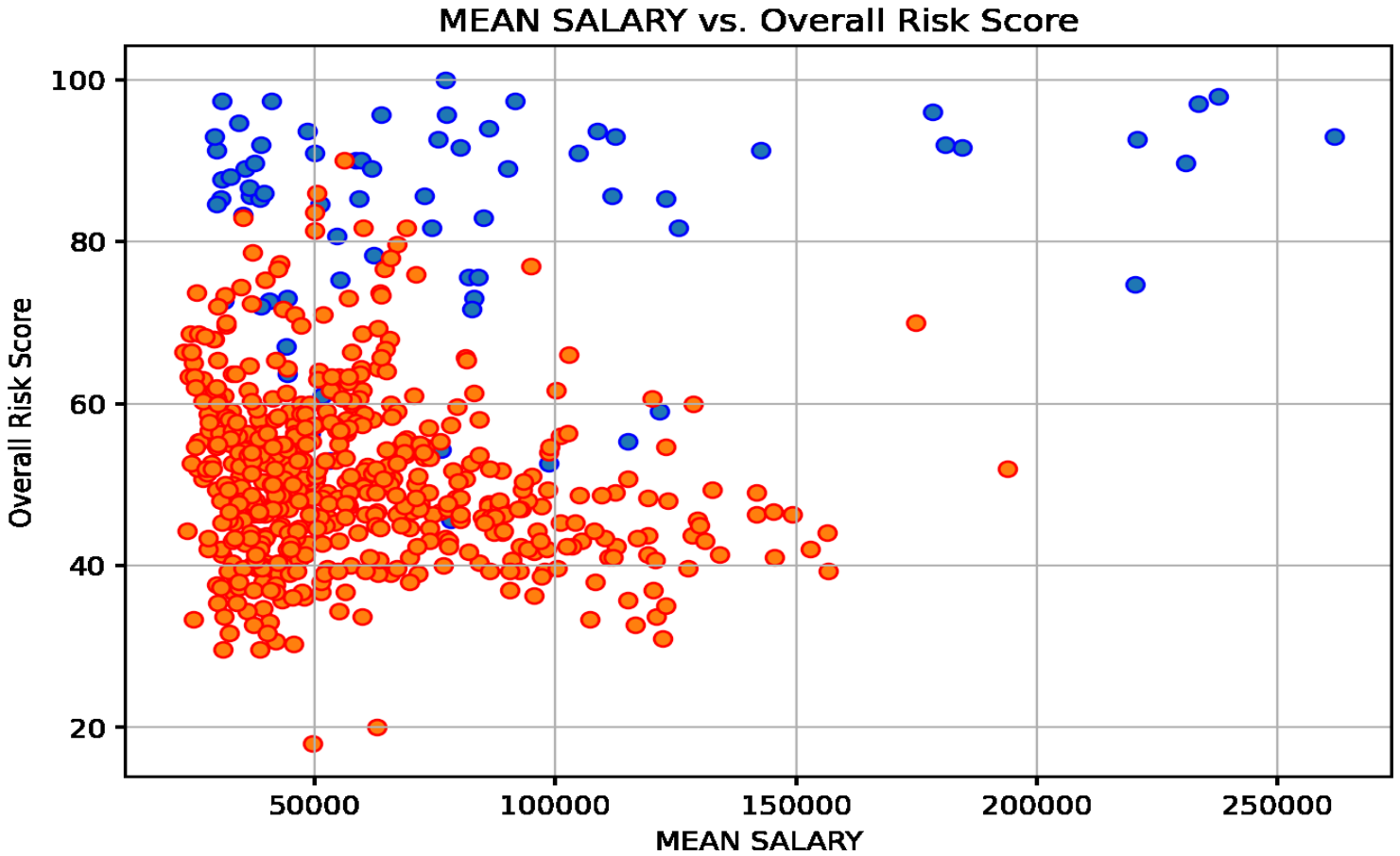
- About 11% of the employable population are in the health profession and on the highest risk score spectrum. However, this concentration in hot spots and hospitals increases the risk for these workers



Health professionals are at highest risk than non-health

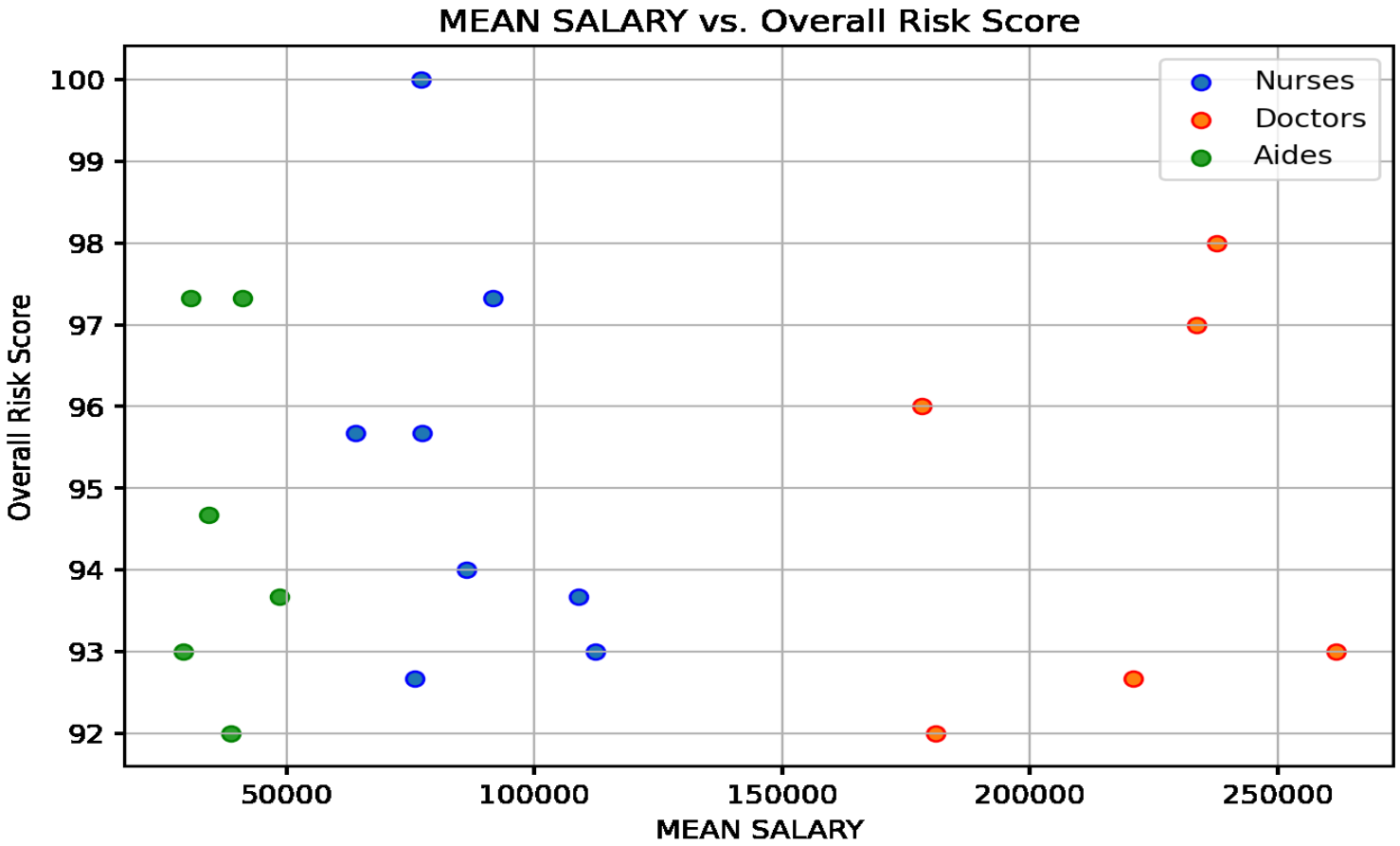


More professions with a lower mean salary in the lower risk spectrum

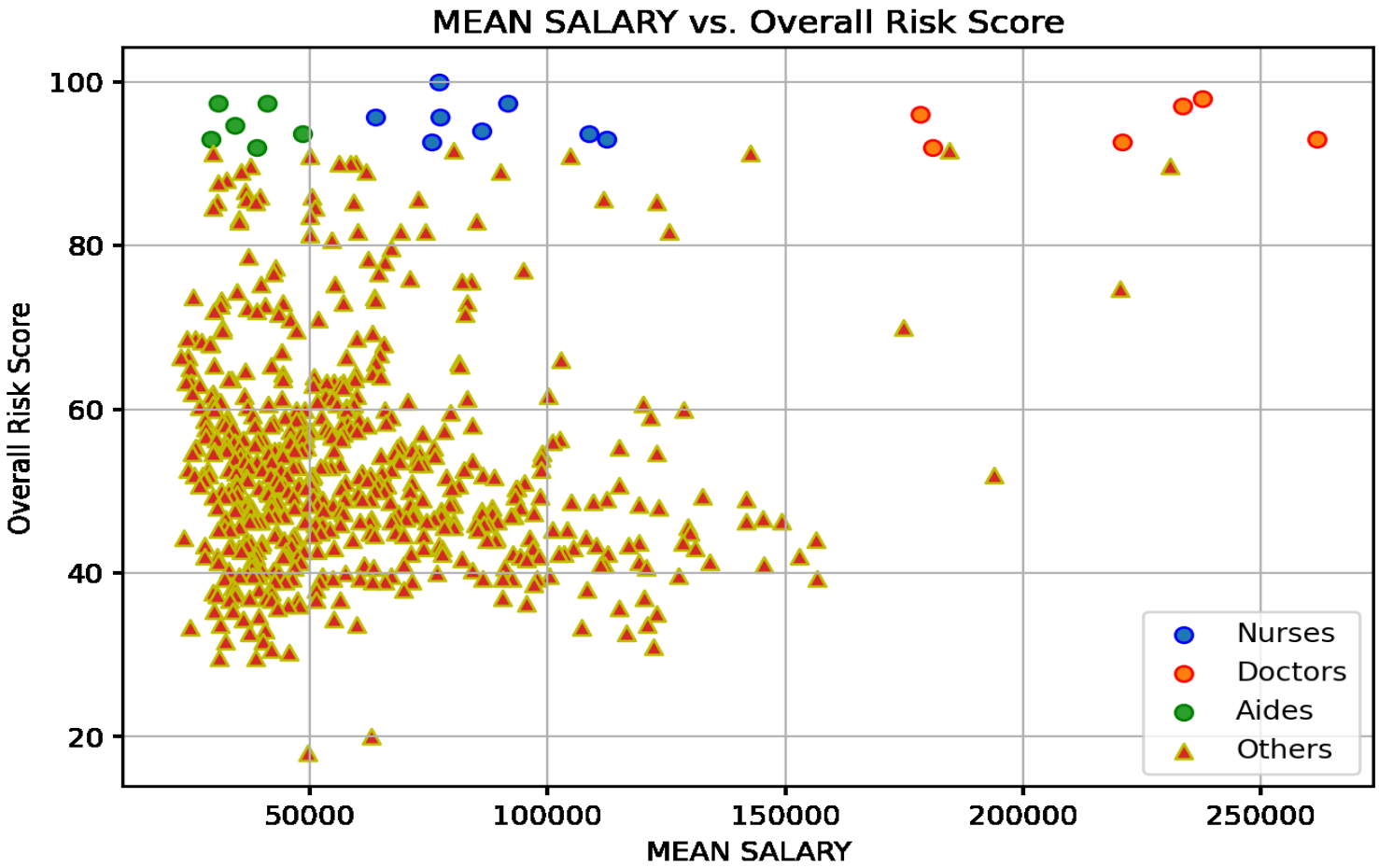


Health professionals are at highest risk than non-health

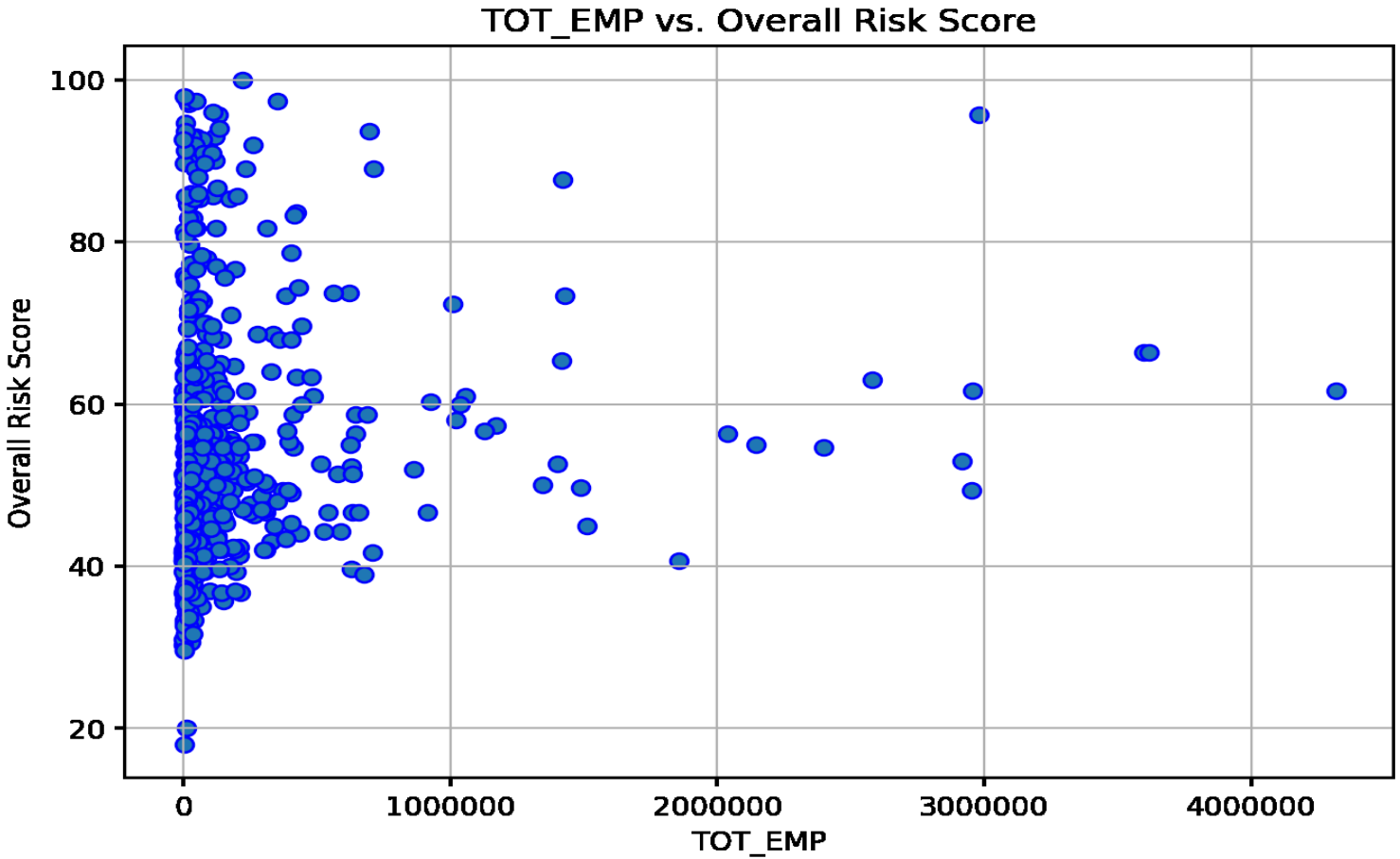
More professions with a lower mean salary in the lower risk spectrum



Top 20 Covid Risk Scores are with the health professionals



Health professionals are at highest risk than non-health and salary does not play a factor



By employment numbers, most of the professions are at lower risk

Summary of Findings

- Health occupations have the highest risk compared to non-health; and salary is not a factor.
- News that many health care workers, especially in the hot spots were at highest risk of contracting Covid is true. Their lack of access of personal protective equipment (ppe) exacerbates their risk
- Doctors are in the same high-risk category as EMT's, even though their salaries can be eight times higher
- Overall, the majority of occupations in the lower salary brackets are also lower risk for Covid
- Top 20 Covid Risk Scores are with the health professionals
- By employment numbers, most of the professions are at lower risk