

Week 3: Identify Risk Factors for Infection

UPDATE *ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE*

Thank you again for the previous analysis. We will next be publishing a public health advisory that warns of specific infection risk factors of which individuals should be aware. Please advise as to which population characteristics are associated with higher infection rates.

Your goal for this notebook will be to identify key potential demographic and economic risk factors for infection by comparing the infected and uninfected populations.

Imports

```
In [ ]: import cudf
import cuml
```

Load Data

Begin by loading the data you've received about week 3 of the outbreak into a cuDF data frame. The data is located at `./data/week3.csv` . For this notebook you will need all columns of the data.

```
In [ ]: ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE
```

Calculate Infection Rates by Employment Code

Convert the `infected` column to type `float32` . For people who are not infected, the float32 `infected` value should be `0.0` , and for infected people it should be `1.0` .

```
In [ ]: ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE
```

Now, produce a list of employment types and their associated **rates** of infection, sorted from highest to lowest rate of infection.

NOTE: The infection **rate** for each employment type should be the percentage of total individuals within an employment type who are infected. Therefore, if employment type "X" has 1000 people, and 10 of them are infected, the infection **rate** would be .01. If employment type "Z" has 10,000 people, and 50 of them are infected, the infection rate would be .005, and would be **lower** than for type "X", even though more people within that employment type were infected.

In []: **ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE**

Finally, read in the employment codes guide from `./data/code_guide.csv` to interpret which employment types are seeing the highest rates of infection.

In []: **ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE**

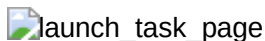
Calculate Infection Rates by Employment Code and Sex

We want to see if there is an effect of `sex` on infection rate, either in addition to `employment` or confounding it. Group by both `employment` and `sex` simultaneously to get the infection rate for the intersection of those categories.

In []: **ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE**

Take the Assessment

After completing the work above, visit the *Launch Section* web page that you used to launch this Jupyter Lab. Scroll down below where you launched Jupyter Lab, and answer the question *Week 3 Assessment*. You can view your overall progress in the assessment by visiting the same *Launch Section* page and clicking on the link to the *Progress* page. On the *Progress* page, if you have successfully answered all the assessment questions, you can click on *Generate Certificate* to receive your certificate in the course.



Optional: Restart the Kernel

If you plan to continue work in other notebooks, please shutdown the kernel.

```
In [ ]: import IPython
        app = IPython.Application.instance()
        app.kernel.do_shutdown(True)
```

ANSWER OR CODE COPYRIGHTED BY NVIDIA DEEP LEARNING INSTITUTE