

# **Analysis of the Information Security Job Landscape in Austria**

---

Markus Reiter

June 5, 2020

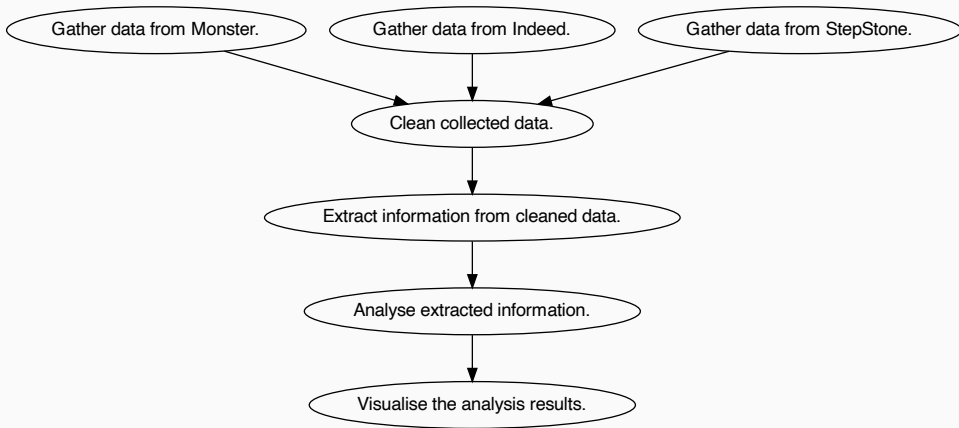
# Outline

- Introduction
- Overview of the Project Workflow
- Evaluation of Scraping Tools
- Implementation
  - Scraping
  - Natural Language Processing
  - Data Analysis
- Findings
- Conclusion

# Introduction

- information security becoming an essential area of expertise
- same understanding of the required skills by both companies and prospective employees
- scraping of three popular job search websites: Monster, Indeed and StepStone
- analysis of scraped data

# Overview of the Project Workflow



## Evaluation of Scraping Tools: OctoParse

- desktop application for scraping
- free tier allows scraping unlimited pages on unlimited computers
- two concurrent scraping jobs
- allows configuring 10 different crawler configurations
- only supported on Windows

## Evaluation of Scraping Tools: ScrapingBot

- does not provide an application at all
- similar to using a proxy server
- HTTP request responds with the contents including any rendered JavaScript
- free plan only includes 100 API calls per month

## Evaluation of Scraping Tools: ScrapeStorm

- desktop application for scraping
- works on Linux, macOS as well as Windows
- free plan allows 10 different crawler configurations
- 1 concurrent local run
- allows 100 exported rows per day

## Evaluation of Scraping Tools: ScrapeStorm Trial

- scrape Monster detail page
- content included in nested `iframe` element
- major problem: extracted content empty
- cannot extract nested `iframe` element



## Evaluation of Scraping Tools: Selenium & Firefox

- supports all of the major web browsers
- only thing needed is the so called “webdriver”
- can be used with a wide array of programming languages
- most importantly: can extract nested `iframe` elements

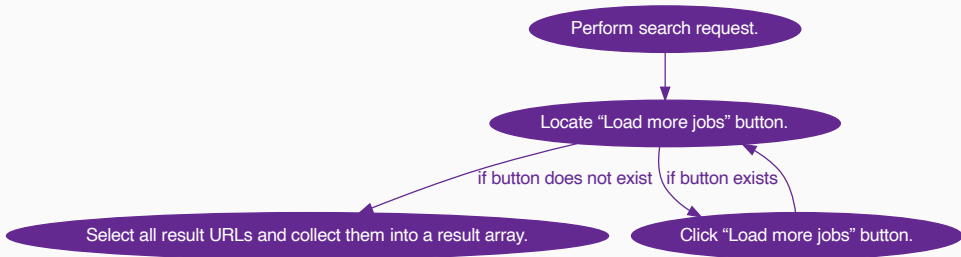
# Implementation

- Scraping
- Natural Language Processing
- Data Analysis

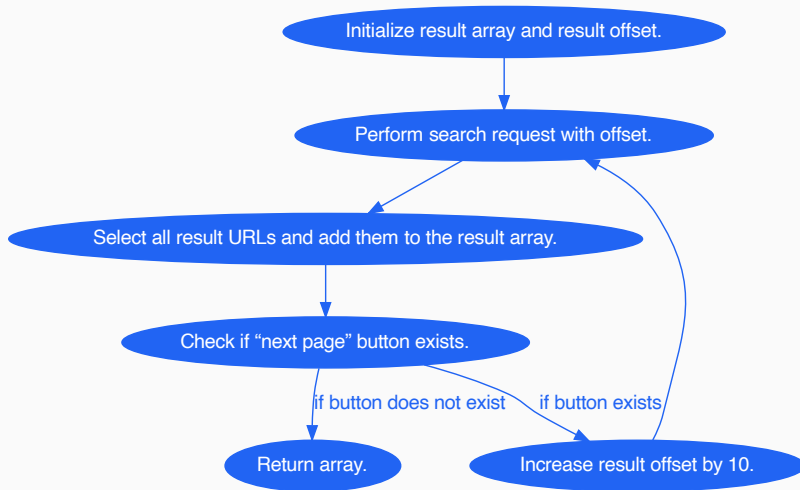
## Implementation: Data Scraping

- scraping each website consists of two main steps
  - perform search request and gather URLs for detail pages
  - scrape data from detail page

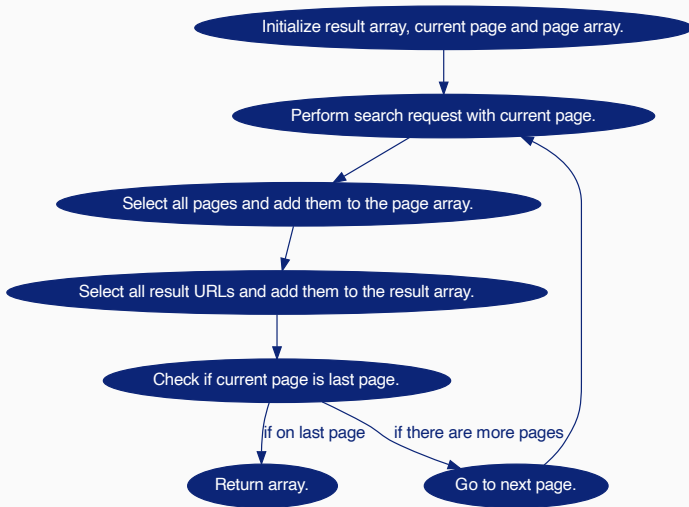
## Implementation: Data Scrapping - Monster URLs



## Implementation: Data Scrapping - Indeed URLs



## Implementation: Data Scrapping - StepStone URLs



## Implementation: Data Scrapping - Detail Page

- title
- metadata: location, contract type
- body content on Monster:
  - find nested `iframe` element
  - use `switch_to` function to switch to the `iframe` element
- body content on Indeed and StepStone:
  - no nested `iframe` element
  - extracted the same way as all other fields

## Implementation: Natural Language Processing

- implemented in Python
- tokenisation using NLTK (Natural Language Toolkit)
- Austrian cities database from [simplemaps.com](https://simplemaps.com)



## Implementation: Natural Language Processing - Extracted Data

- average salary
- location (including mapping cities to states)
- employment type (full-time vs. part-time and permanent vs. temporary)
- required education (Matura, bachelor's degree, master's degree etc.)
- required certifications

## Implementation: Natural Language Processing - Example Salary

|              |              |             |
|--------------|--------------|-------------|
| 43,595 Euros | EUR 2.600,00 | 65,000 €    |
| 30k Euros    | 70.000 €     | €60,000     |
| 3,100.00 EUR | € 4.209,-    | €85k        |
| EUR 2.800,-  | € 3.933,0    | €3.007,20   |
| EUR 50k      | € 55.000,-   | 80.000 Euro |
| EUR 5700     | € 65.000,—   | 60000 €     |

- needed to be normalised to XXXXX.XX € format
- extracted using 2-grams with number and € symbol

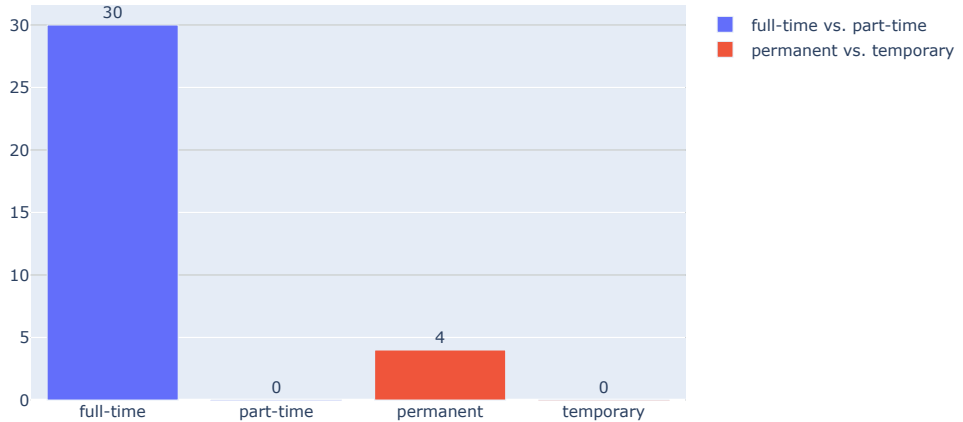
## Implementation: Data Analysis

- Jupyter Notebook (interactive Python environment)
  - very useful for quickly iterating over revisions
  - graphs are displayed in the same environment as the code
- pandas framework for manipulating data
- Plotly for generating graphs

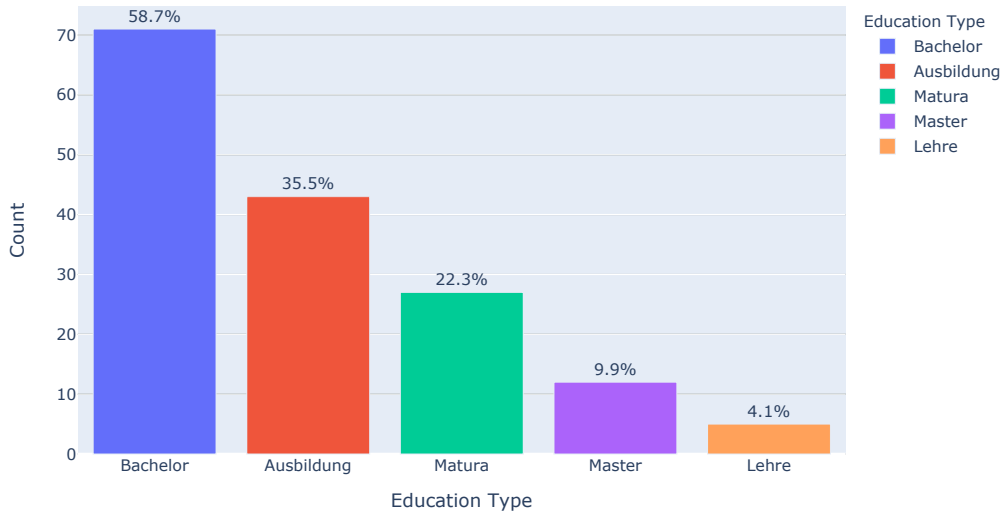
# Findings

- search term “information security”
- 1538 scraped jobs in total
  - 993 from Monster
  - 282 from Indeed
  - 263 from StepStone
- 121 actual information security jobs
  - filtered according to job title

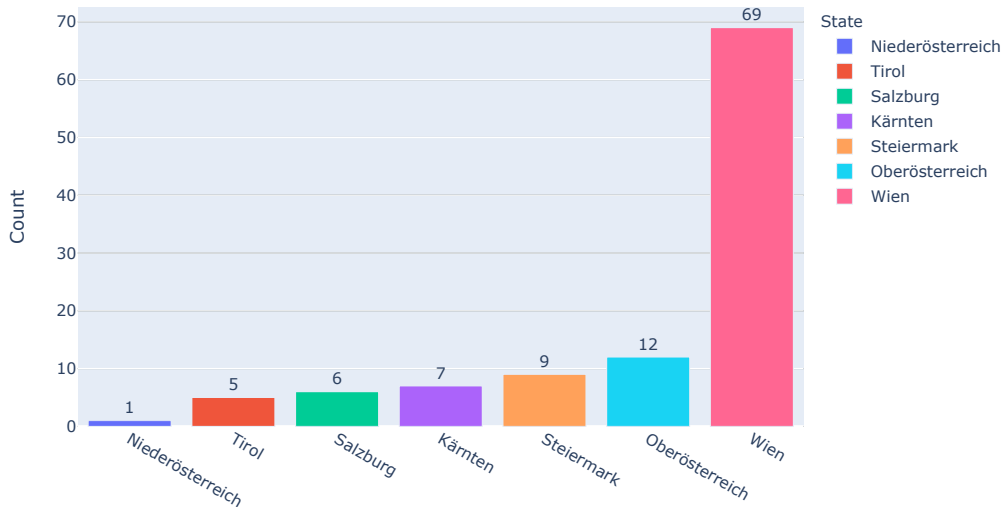
## Findings: Employment Type



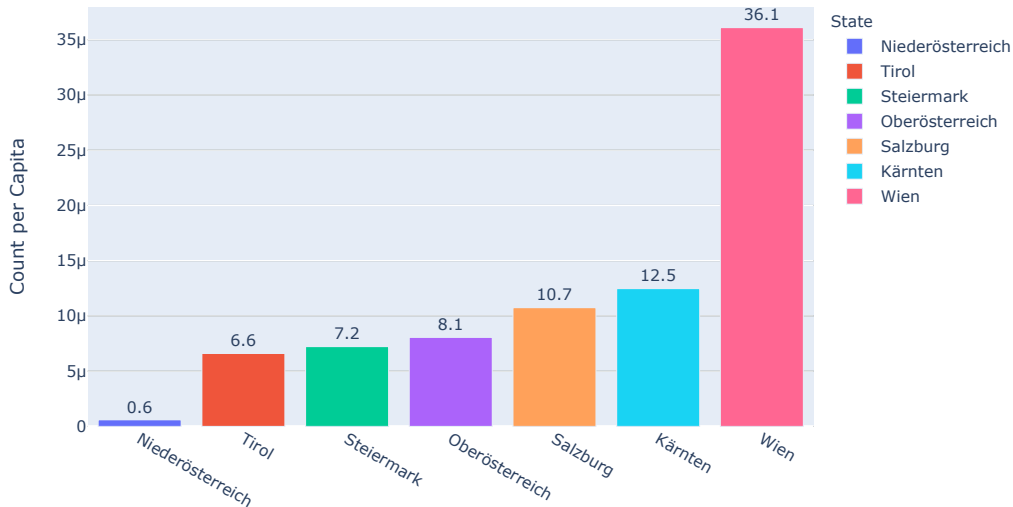
## Findings: Education Type



## Findings: Location

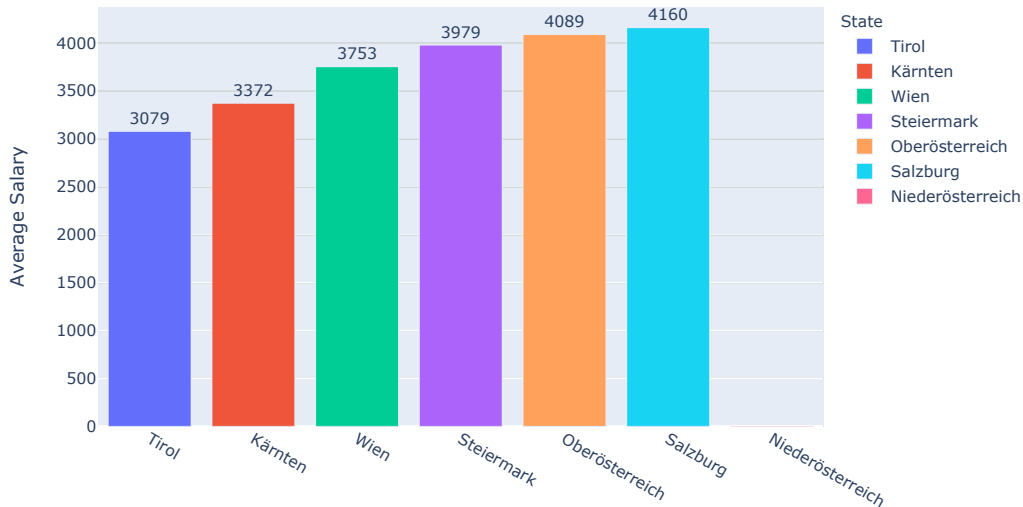


## Findings: Location per Capita

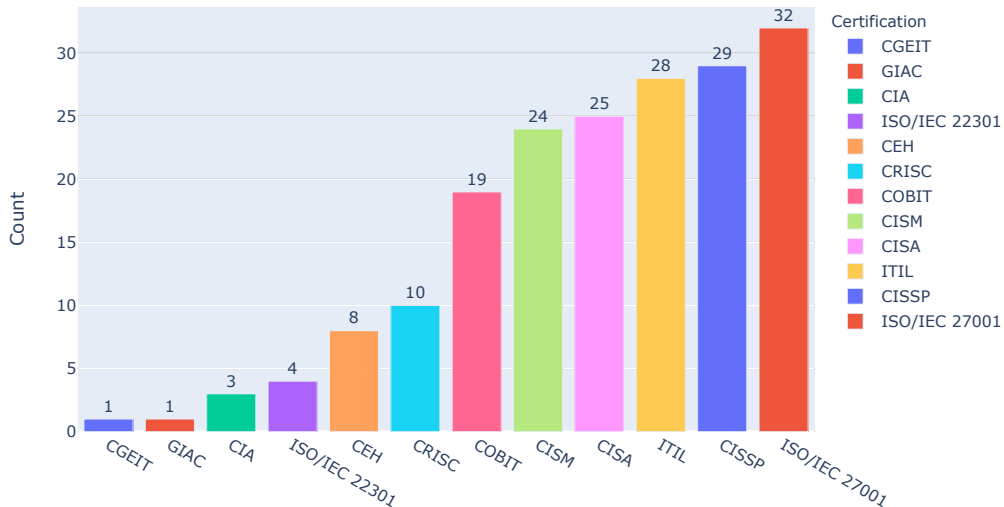




## Findings: Average Salary



## Findings: Certifications



## Conclusion

- compared multiple scraping tools
- developed a program for scraping data from three job search sites
- findings can be used to inform new decisions on what should be taught in universities
- more data points could be implemented in a future revision of this survey
- survey could be conducted again within a longer time span
- survey could be conducted again with additional job search websites