

## **Documentation of the Workflow**

*by Denin Lunja (2729903), Maiss Aldahoul (2783572), Nicola Roos (2869010), Rasmus Laubinger (2868328) and Tamara van Maarseveen (2767944)*

Lecturer: Dr. Lorella Viola

Vrije Universiteit Amsterdam

Introduction to Digital Humanities and Social Analytics

Group 1

Word count: 2978

Date: 15.10.2024

## 1. Project Overview:

### Group topic:

The Dutch East India Company's 18th Century Workforce.

### Data description:

The dataset consists of three files:

- VOC pay books (VOC-soldijboeken).
- VOC passenger lists (VOC-opvarenden).
- VOC beneficiaries (VOC-begunstigden).

The original dataset had significant weaknesses, limiting research validity, which Petram et al., (2024) addressed by disambiguating individual records, standardizing names, and adding wage and rank details, making it possible to study the VOC workforce from various perspectives with a level of reliability and validity that was not previously achievable.

### Research Question:

*How did the regional origins of workers within Europe impact their work conditions and opportunities in the VOC during the 18th century?*

### Thesis Statement:

Our original thesis statement was as following;

*Despite the VOC's multi-national workforce, analysis of 18th century pay ledgers will illuminate the professional advantages awarded to Dutch-originating employees, with non-Dutch-originating employees having been over-represented in lower-paying jobs and less likely to receive promotions.*

However, due to the complexity of the provided dataset, we could not address reemployment cases. The dataset primarily focuses on employment records rather than individual workers. The original VOC pay ledgers, designed solely for salary calculations, closed employment records at the end of an employee's service. Subsequent re-employment was recorded separately, leading to fragmented data for those who served multiple times. Thus, we had to rethink our research statement and consider a more focused one.

Our new thesis statement:

*While previous research has treated the VOC's European workforce as a monolithic group, an examination of VOC pay ledgers, passenger lists, and beneficiaries demonstrates that this workforce was actually diverse and heterogeneous, with significant variations in work opportunities and conditions.*

### **Overview:**

The Dutch East India Company (VOC), established in 1602, was a robust trading enterprise that dominated long-distance trade between Asia and Europe until the late 17th century. The company aimed to control the Indian Ocean trade routes and surpass local traders in Asia (Gaastra, 1991). To consolidate its influence, the VOC employed individuals from diverse backgrounds, often using this as a tactic to manage and control local labor, reinforcing colonial hierarchies and power structures.

When examining modern workplaces, it is essential to recognize how historical colonial systems have influenced today's workforce dynamics. Significant parallels exist between the VOC's practices and current discussions on diversity, access to economic opportunities, and organizational hierarchies. This historical context sheds light on the persistent marginalization of certain groups in the workforce, emphasizing the urgent need to address these disparities and foster more equitable working environments.

Previous research has shown that Dutch workers had higher promotion rates than their non-Dutch counterparts (Bruijn & Gaastra, 2012; Wezel & Ruef, 2024), and Europeans, in general, enjoyed better working conditions than their Asian colleagues (Lucassen, 2004). This study aims to extend that analysis by examining the job prospects within the European workforce, which was less homogenous than often assumed.

The findings reveal that Dutch workers had significantly higher chances of promotion than non-Dutch employees, underscoring a notable gap in job advancement opportunities. Moreover, Europeans were offered much better working conditions than their Asian counterparts, indicating an apparent inequality that favored European laborers. The limited documentation of Asian workers within the company, however, highlights an ongoing issue of historical oversight. This oversight not only reflects the unequal treatment of non-European workers but also impedes a comprehensive understanding of the VOC's labor dynamics.

## **2. Data Acquisition**

The dataset for this study consists of three files: The VOC pay books (VOC-soldijboeken), VOC passenger lists (VOC-opvarenden), and VOC beneficiaries (VOC-begunstigden). These records offer valuable insights into the workforce's origins, ranks, and wages, which functions as a resource for studying the company's operations. Petram al. (2024) enriched the dataset through several key steps of analysis. This includes standardizing the names of places to reduce inconsistencies across the records, clustering name variations, and addressing ambiguities in job titles. Despite the extensive records, the data still has its limitations. Firstly, the group sizes vary significantly, which may cause results to be skewed due to the disproportionate representation of workforce segments. Furthermore, the documentation mainly focuses on European workers, as it only includes records from the ships that departed from one of the six chambers in the Netherlands. Previous research indicates that a large portion of the VOC workforce was composed of Asian workers and slaves, who are underrepresented in this dataset, limiting the scope of our analysis. Another challenge is the inclusion of workers with multiple contracts under different job titles, treated as separate cases in the analysis. Furthermore, rank classification poses particular difficulties due to ambiguous job titles and shifts in rank meanings after 1784, when higher designations were assigned to lower roles to attract skilled laborers (Lucassen, 2004).

## **3. Methodology**

For our methodology, we first transferred the datasets (voc\_ranks, voc\_places, and voc\_persons\_contracts) from Excel into the statistical analysis software SPSS. These datasets were then merged into one detailed dataset for analysis. Next, we performed data cleaning by deleting cases without a specified place of origin (4703 cases) and those whose place of origin was not standardized or could not be affiliated with any known region (132,589 cases). In total, 137,292 cases were removed, leaving 636,907 valid cases for further analysis. We decided to retain cases with multiple entries, as reemployment data is crucial for understanding work opportunities within each region. Statistically, given the size of the dataset, all regions include cases with multiple entries, and therefore this does not significantly skew the data for our analysis.

#### **4. Workflow steps**

Our research began by defining clear objectives and formulating our research question, focusing on how the place of origin impacts variables such as rank and wage. After understanding the dataset, we reviewed the existing preprocessing steps to ensure the data was properly enriched and ready for analysis. A comprehensive literature review was conducted to explore previous studies, allowing us to identify relevant material. Based on this, we selected key variables, including place of origin as the independent variable and rank and wage as dependent variables, to guide our analysis. Therefore, the workflow by steps was as follows:

1. Defining the objectives and the RQ
2. Understanding the dataset
3. Learning about the background of the topic
4. Deciding what we want to focus on in our research
5. Statistical analysis
6. Discussion of results within the team
7. Visualization of the data
8. Drawing conclusions

#### **5. Challenges & Solutions**

The majority of challenges we encountered were related to the dataset:

The sizes of the origin groups vary significantly, which may lead to skewed results. Further, the documentation primarily focuses on European workers, despite Asian workers (including slaves) having made up a substantial portion of the VOC workforce. The dataset is limited to records from ships that departed from one of the six chambers in the Netherlands, narrowing our study to a subset of the overall VOC workforce. It also primarily focuses on employment records rather than individual workers. The original VOC pay ledgers, designed solely for salary calculations, closed employment records at the end of an employee's service. Subsequent re-employment was recorded separately, leading to fragmented data for those who served multiple times. For these challenges, the only available solution was to be aware of them, consider them when designing the methodology and include them in our discussion.

Another challenge was that rank classification is particularly prone to error due to ambiguous job titles and changes in rank meanings after 1784, when higher designations were applied to lower roles.

In order to deal with the ambiguous job titles and be able to categorize employees according to job ranks, we coded job titles into 3 groups: “Senior Rank”, “Rank”, and “Junior Rank”. The grouping was based on median wage. Jobs that did not have “senior” or “junior” in the rank title were sorted into the group category “Rank”.

## **6. Ethical Considerations**

The data used to conduct our research underwent multiple steps. The original physical pay-ledgers and documents kept by the VOC were digitized and fully published in 2012 (Petram et al., 2024). This collection is hosted by the Dutch National Archives (cf. VOC: Opvarenden) and is openly accessible. Petram et al. (2024), the authors of the enriched dataset at the base of our study, not only published their dataset and code with open access, but encouraged its use for further research.

While the subset of data we analyzed does include personal and potentially identifiable data of individuals, such as their full names and places of origin, we consider our research in line with ethical standards for several reasons: Firstly, the afore-mentioned personal data has been publically accessible for over a decade now. Secondly, our analysis and results are aggregated, meaning that individual identities are not in focus and cannot be traced back to from our results. Thirdly, it can be said with certainty that the persons included in the dataset have been deceased for at least a century. While there is ethical debate around the deceased’s right to privacy, we have taken this matter into consideration. We feel that the chance of our research doing harm to any person is negligible. We could not find reasonable support for the possibility of our research exposing any individual or otherwise rendering them more vulnerable to scrutiny than the original and enriched datasets did.

Further, there is no particularly sensitive data (such as sexual orientation, religion) contained in our dataset.

Ethical considerations around bias are best split into two areas: Bias already contained in the dataset before our processing, and bias added after.

The former was addressed in Petram et al.’s (2024) introduction of their dataset, and detailed once more in our documentation. The main concerns were around Dutch-centric recording of VOC employees’ information, resulting in less workable data for non-Dutch-originating employees. Unfortunately, counteracting measures such as verifying non-Dutch places of origin or integrating other sources in an attempt to reliably reconstruct missing or ambiguous information were not feasible within the scope of our research project.

Instead, we opted to acknowledge, evaluate and document this bias and its possible implications.

In order to mitigate our own bias, we familiarized ourselves with the intricacies of the enriched dataset, researched related topics and carefully considered each decision that stood before us. Any concerns formed throughout the process of the research were voiced and investigated, prompting changes in methodology if warranted. Any potential bias we could identify has been thoroughly documented.

## **7. Results**

The first part of the analysis showed the unequal distribution of the workers within Europe. Especially the dominance of workers originating in the Dutch region (62.3%) confirmed what Petram et al. (2024) already described. Firstly, when looking at the ranks, the findings showed that most workers worked as sailors (50.8%), followed by soldiers (27.4%) and craftsmen (4.4%). Yet, substantial differences can be observed between the regions of origin. What stands out are Scandinavian workers (75.7%) and workers from the Iberian Peninsula (76.4%), who had an unproportionally high share of sailors. A possible explanation could be that those regions were historically experienced in seafare and were possibly more qualified to work as sailors. Opposingly to this stand the German interior lands and France, who are both low on sailors but have a high share of soldiers (65%; 59.7%).

In regards to the order of ranks, the data exhibits that 50% of workers worked in junior (low wage) positions, 45.2% in medium positions and merely 4.8% in senior ranks. There was a noticeably high proportion of senior ranked Dutch workers (6%) compared to the other regions that were all below the overall average of 4.8% senior ranks. Yet, despite this high proportion of Dutch senior ranked workers, the average rank of all workers was equal to the overall average ( $M = 1.55$ ). Instead, Scandinavian workers had the highest average rank ( $M = 1.83$ ), opposing the German interior lands that were ranked lowest on average ( $M = 1.26$ ).

Surprisingly, throughout the analysis we discovered a negative correlation between the mean rank and contract length. This suggests that on average, workers in lower ranked positions had to work on longer journeys compared to higher ranked positions.

The analysis of wages offered an interesting finding. Thus, workers from outside of Europe showed in fact the highest average wage of 13.13 guilders per month followed by

Dutch ( $M = 17.87$ ) and Scandinavian ( $M = 12.08$ ). France ( $M = 10.44$ ) and German interior lands ( $M = 10.52$ ) portrayed the lowest average wage.

Finally, our data showed that the most common reasons for the end of contract were Deceased (46.3%) or Repatriated (30.7%). This shows how dangerous the seafare at that time still was, but also the economic prospect that people saw in it for them to still be willing to accept that high risk of dying.

Concludingly, the results show that the VOC's European workforce cannot be considered a homogenous entity but a diverse group with varying wages, ranks and contract details. Therefore, future research should consider these regional differences and not treat them too simplistically. Additionally, our study only found limited support for previous research stating that Dutch workers experienced better working conditions compared to non-Dutch workers (Wezel et al., 2024). While we found that Dutch employees had the highest share of senior ranks, they did not stand out regarding average wages or average ranks.

Returning to the research question, *How did the regional origins of workers within Europe impact their work conditions and opportunities in the VOC during the 18th century?*, our findings show that VOC employees' regional origins did in multiple cases correlate with factors that affected their work conditions, such as their type of job and their level of rank. This suggests that differentiating between various European origins when conducting research on the VOC's workforce will provide insight otherwise lost to generalization. However, the results do not permit a definitive confirmation that regional origins *impacted* working conditions and opportunities, as no causation could be proven. In summary, our research cannot fully answer the research question, but uncovered points of interest which could help guide future inquiries.

## **8. Documentation & Sustainability**

Project data: <https://github.com/DeninLunja/IDHSA-Team1>

The project's data includes the combined VOC datasets on SPSS, the output of the analysis and the syntax of the analysis steps. Reviewing the syntax and the output helps in increasing transparency and replicability. The syntax, for instance, clearly shows the analysis steps that have been taken to acquire the presented results, including the detailed documentation of specific decisions taken throughout the analysis. This ensures the



opportunity for an objective review of the entire study by other researchers. Additionally, the combined datasets in SPSS can be further used to examine different aspects of the VOC workforce. Whereas this study focused on comparing different regions, the data can also offer insights into employees' relationships through their monthly letters, or examine reemployment and promotion rates. The separate VOC datasets were merged into a single dataset whose data was already pre-processed. It is therefore very well-suited for further analysis on the VOC.

## **9. Reflection**

Successfully delivering a workflow was a highly collaborative effort. The nature of the dataset required the application of various skills in order to conduct a meaningful analysis: The size of it demanded a comprehensive understanding of statistics and proficiency in a statistics software, its historical and societally consequential topic warranted sociological sensitivity and media literacy, the digital aspect called for fluency in digital methods and upkeep. These skills were represented across research members, making it imperative that we be able to communicate with each other, unify our knowledge and logically think of next steps. Finally, all of these aspects needed to be communicated effectively and accessibly to the broader public.

It is due to our shared understanding of the complexity of the task, willingness to change direction and trust in each other's competencies that we feel the workflow of our project proceeded quite seamlessly. The direction of the project was re-evaluated routinely, based on the outcomes of each newly taken step and its implications for what was feasible. A tentative direction or goal was agreed on for the next step, and the process subsequently repeated. The process was not linear, as each re-evaluation included verifying that previous choices were still in line with the research goal and adjusting prior choices when needed. However, the frequency of these check-ins seemed to keep adjustments small and relatively unintrusive. For example, after writing a team charter, it was agreed that a research question would not be formed until we had a rudimentary understanding of the dataset and familiarized ourselves with relevant literature. Following this, we centrally agreed on a preliminary research question, to be refined after more literature research and achieving a deeper understanding of the dataset. As prophesied, at the next meeting, we deemed it necessary to tweak the research question. This anecdote is representative of much of the workflow. The remaining two big changes were excluding re-employment rate from our analysis due to feasibility issues, and updating our thesis statement to reflect our findings.

The resulting findings lay the groundwork for a wide array of future research. Parallel to the limitations we experienced, an in-depth exploration of the working conditions for Asian employees as well as analyses of re-employment and promotion rates could prove insightful. The change in workforce composition over time could be examined, along with its correlations with (Dutch) public opinion on the VOC. Drawing from our findings that there was a strong correlation between certain places of origin and certain job titles, one could investigate whether recruitment strategies and advertisements differed between regions. Finally, one could consider a comparison with patterns of multinational corporations today, such as the outsourcing of dangerous, low-prestige jobs to foreign workers.

Overall, the workflow was very effective. The strategies discussed above guided the research well. Based on our experience, approaching a digital humanities project with meticulous consideration, frequent re-evaluations and a shared awareness of the iterative research process will render us well-equipped to grapple with a given project's idiosyncrasies.

## References

- Bruijn, J.R. & Gaastra, F. S. (2012). The career ladder to the top of the Dutch East India Company: Could foreigners also become commanders and junior merchants? In M. van der Linden & L. Lucassen (Eds.), *Working on Labor* (pp. 213-235). Brill.
- Lucassen, J. (2004). A Multinational and its Labor Force: The Dutch East India Company, 1595–1795. *International Labor and Working Class History*, 66(66), 12–39.  
<https://doi.org/10.1017/S0147547904000158>
- Petram, L., Koolen, M., Wevers, M., & van Lottum, J. (2024). Charting Lives and Careers: Enriched Data About the Dutch East India Company's Eighteenth-Century European Workforce. *Journal of Open Humanities Data*, 10, 44–44.  
<https://doi.org/10.5334/johd.210>
- VOC: Opvarenden. (n.d.). Nationaal Archief.  
<https://www.nationaalarchief.nl/onderzoeken/index/nt00444?activeTab=nt>
- Wezel, F. C., & Ruef, M. (2024). Cracking the deck: National origins and promotions in the Dutch East India Company, 1700–1796. *Organization Studies*, 45(7), 1039–1061.  
<https://doi.org/10.1177/01708406241248985>

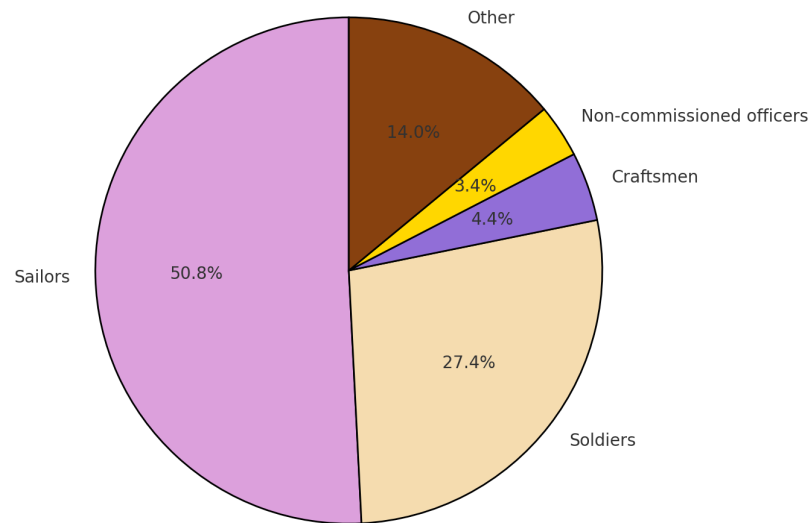
**Appendix A**  
**Visualizations of data**

**Table 1: Regional Origins Distribution**

<b>Region</b>	
Dutch Republic and Dutch-speaking area of southern Netherland	62.3 %
Low German-speaking area	16.6 %
Scandinavia	8.8 %
German interior lands	8.6 %
British Isles	0.4 %
France	1.6 %
Iberian Peninsula	0.2 %
Italy and Corsica	0.5%
Eastern and Southeastern Europe	0.2 %
Outside of Europe	0.9%

*T1: Regional Origins Distribution. Table containing the percentage of the overall workforce that originated from each specified region of origin.*

**Pie Chart 1: Ranks**



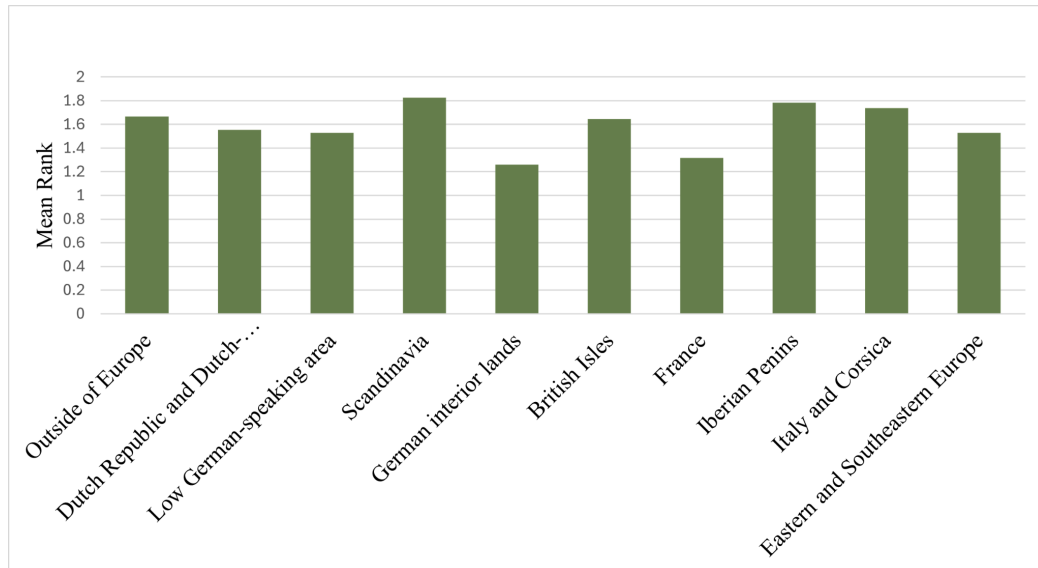
*PCI: Ranks. Shows the overall distribution of jobs within the VOC workforce, represented as broader categories, i.e. Sailors (pink), Soldiers (beige), Craftsmen (purple), Non-commissioned officers (yellow) and Other (brown).*

**Table 2: Rank Distribution**

Region	Junior Rank	Rank	Senior Rank
Dutch Republic and Dutch-speaking area of southern Netherland	50.6%	43.5%	<b>6%</b>
Low German-speaking area	50.1%	46.8%	3.1%
Scandinavia	20.7%	75.8%	3.5%
German interior lands	<b>75.7%</b>	22.4%	1.9%
British Isles	38%	59.6%	2.4%
France	<b>70%</b>	28.3%	<b>1.7%</b>
Iberian Peninsula	23.2%	75.2%	<b>1.6%</b>
Italy and Corsica	28.3%	69.5%	2.2%
Eastern and Southeastern Europe	49%	49%	2%
Outside of Europe	37.2%	58.9%	3.9%
<b>Total</b>	<b>50%</b>	<b>45.2%</b>	<b>4.8%</b>

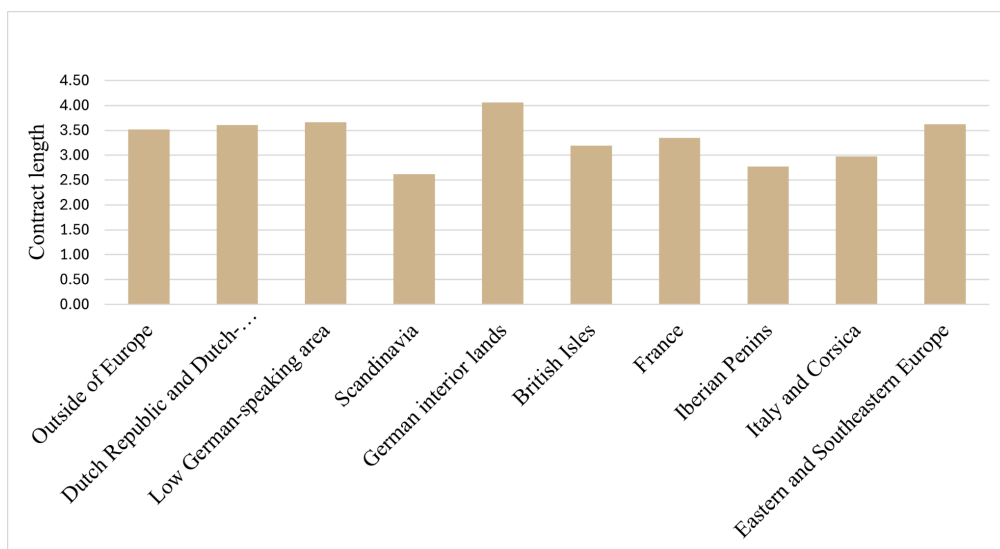
*T2: Rank Distribution. Table containing the percentages of employees per regional origin (column 1) who held a junior rank (column 2), medium rank (column 3), and senior rank (column 4).*

**Bar Graph 1: Mean Rank**



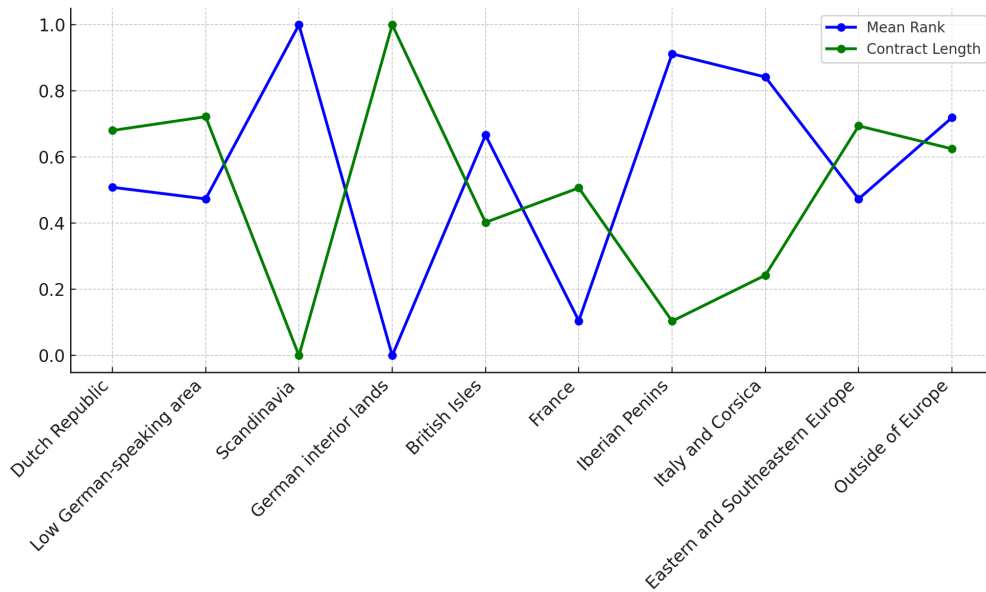
*BG1: Mean Rank. Shows the mean rank (represented on the y-axis with 0 = junior rank, 1 = medium rank, 2 = senior rank) of VOC employees based on their regional origins (x-axis).*

**Bar Graph 2: Contract Length**



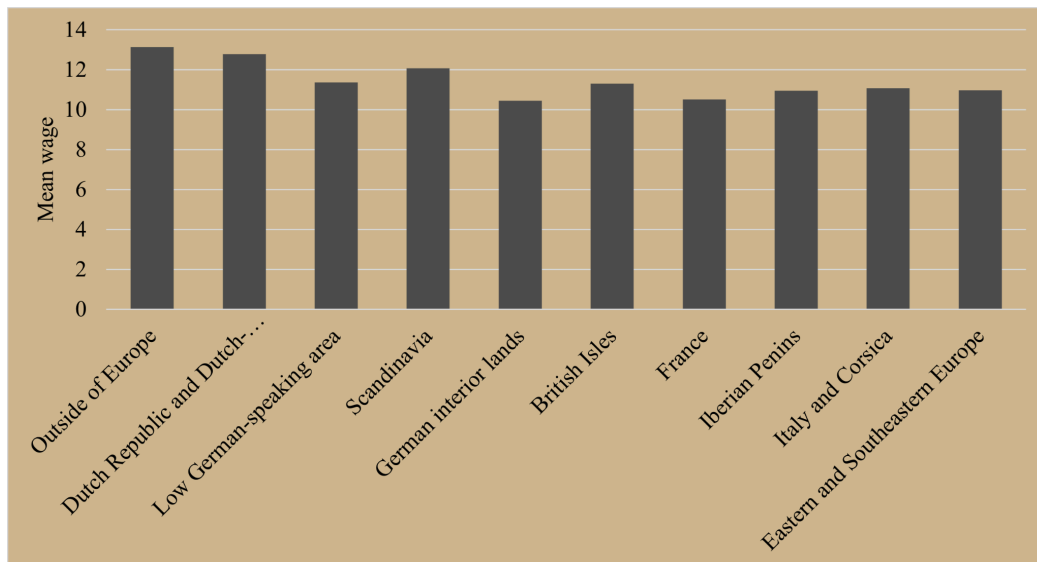
*BG2: Contract Length. Shows the mean contract length in years plotted on the y-axis for VOC employees based on their regional origins (x-axis).*

**Line Graph 1: Mean Rank x Contract Length**



*LG1: Mean Rank x Contract Length. Depicts a negative correlation between mean rank (blue) and contract length (green), with values standardized to values between 0-1 (y-axis), and sorted by employees' regional origins (x-axis).*

**Bar Graph 3: Mean Wage**



*BG3: Mean Wage. Shows the mean wage in guilders per month (y-axis) earned by VOC employees sorted into groups of regional origin (x-axis).*