Questions for the Industrial Master's Programme in Swedbank

Name: Xinjian Zhang
Program: MSc Software Engineering
Email: xinjian.zhang@ut.ee

Swedbank

Preferred topic: 1. Develop advanced methodology for creating synthetic data representative of our customers' behaviors

---

1. Why do you want to work at Swedbank?

- **Commitment to Sustainable Economy:** Swedbank's focus on supporting a healthy and sustainable economy resonates with my personal and professional values. The bank's proactive approach to using data science in financial services to promote societal growth is inspiring.
- **Data Science in AFC:** Working in Swedbank's Anti-Financial Crime (AFC) department is a great fit for my goal of using data science to make a big difference in society. I am eager to contribute to the development of solutions that prevent financial crimes by leveraging my expertise in complex data analysis and modelling.
- **Learning and Growth:** Swedbank's environment is distinguished by its use of the SAFe methodology and its focus on collaboration and stakeholder engagement. I am particularly interested in the bank's innovative projects, closely related to my academic and project activities.

2. Who or what inspired you to work in data science?

My interest in the stories hidden within numbers, as well as the potential to use those stories to solve real-world problems, inspired me to pursue a career in data science. My interest grew during my university studies, where I took courses in statistics, programming, and business and product-related analysis and management. The challenge of making sense of complex data sets, as well as the satisfaction of discovering insights that could lead to actionable environmental strategies, deeply motivated me. It made me realise the value of data science, which benefits business optimisation and positive societal transformation. This realisation has been my driving force, guiding me toward a career that I am passionate about.

3. Describe your favorite data science project that you have been involved in, focusing on why the project was important and the tangible outcomes your work led to.

The project, `**Research on the Urgency of Carbon Emission Issues from Sewage Plants**`, is the data science project that I favourite. The goal of this project was to analyse carbon emission trends from sewage plants in order to inform better environmental policy decisions.

- **Objective:** The goal was to model the carbon emission trends of sewage treatment plants across the country to predict when emissions would peak and to assess the urgency of emission reduction efforts.

- **Data Collection and Preprocessing:**
    - Utilised Python's Pandas library to aggregate and clean extensive datasets from various provinces, focusing on COD (Chemical Oxygen Demand) and direct and indirect carbon emission figures over the past decade.
    - Applied Numpy for efficient numerical analysis of the cleaned data, preparing it for further modelling and analysis.

- **Analysis and Modeling:**
    - Implemented time series analysis using Python's Statsmodels package to understand emission trends and forecast future emission peaks. This was crucial for determining the timeline for potential carbon neutrality in sewage treatment efforts.
    - Developed mathematical models in MATLAB to analyse the relationship between COD levels and carbon emissions, providing a deeper understanding of how sewage treatment processes contribute to overall emissions.
    - Employed Scikit-learn for machine learning models to predict emission levels based on a range of inputs, including treatment techniques, plant capacities, and regional factors. This helped identify key drivers of emissions and potential areas for intervention.

- **Visualisation:**
    - Used Matplotlib and Seaborn for Python to create visualisations that communicated our findings to non-technical stakeholders, illustrating trends and predictions that underpinned our recommendations for policy and practice changes.

The project was an extensive application of data science techniques to tackle environmental issues. It highlighted the critical role of data preprocessing and the significant insights derived from statistical analysis.

4. What do you want to achieve and learn working with Swedbank during your industrial masters?

- **Contribute**: I will use my data science expertise to support the team and learn academic skills and comprehensive abilities from the best in the industry.
- **Expand Skills**: I look forward to deepening my knowledge of Python and other tools, embracing best practices in data science, and understanding the intricacies of financial data. I'm passionate about using data science to solve cross-disciplinary challenges.
- **Collaborate and Learn**: Collaborate closely with Swedbank's data science team and stakeholders, drawing on their experiences and insights. This collaborative environment will be critical to my professional growth and understanding of the financial industry's challenges and opportunities.

Last words

With my skills in Python, R, and Matlab and other programming experiences, I am excited to contribute to the financial industry. This program is critical to my career aspirations, and I look forward to making meaningful contributions.