Topic: Expansion of the remote labour market post covid-19

1. **What is your research question?**

This project investigates the *expansion/development* of remote work in the Danish labour market in recent years (xxxx-2020), especially since the outbreak of covid-19 in March 2020, and cautiously aims to predict the scope of remote work in the future within different industries.

1. **How are you going to find out/what methods are you planning to use?**

The first part of the project consists of a descriptive analysis of the trend in job postings that include the option of working remotely both in absolute numbers but also relative to the total number of job postings. This is done merely to confirm (at least superficially) the trend that many media outlets and academics have foreseen whereby the option of working remotely is becoming increasingly available.

we will first need scrape the job portal [www.jobindex.dk](http://www.jobindex.dk) for all current and archived job postings. We will then define some keywords (or expressions) such as “hjemmearbejde, hjemmekontor, arbejde hjemmefra, etc” that we expect job postings where working remotely is allowed (or encouraged) would contain. We will then search for this expressions in or job data and classify jobs into remote/not remote (we could also give them a score since some jobs allow for more remote work than others do? ).

We hope to classify the job postings into job functions (preferable DISCO-08 codes), industries (preferably DB07 codes) and geographical areas to identify the job types, industries and geographical locations in the country where remote working is especially expanding.

The descriptive analysis will map the historic and current state of working remotely in Denmark and provide a stepping stone for the second part of the project, which aims to identify the scope of remote work in the future by training a machine learning model to be able to estimate the probability that job can be performed remotely, given the features described in a job posting . These probabilities can be interpreted as a proxy of the level of flexibility each occupation has in terms of working remotely - one could for example expect that jobs within the service and hospitality industry cannot be fulfilled remotely. The prediction can help reveal the potential of remote work in the future.

We are aware that using data from job openings means that we are focusing on new jobs/vacancies and we cannot say that the same trends apply to jobs that were already taken. However, it seems reasonable to assume that this is the case.

**Challenges:**

* Jobindex categories do not convert directly into either DISCO-08 or DB07
* The current state of the economy
* HOW TO DO THE ML STUFF?

1. **Why is this going to be interesting to read?**

The spread of CO-VID 19 has forced governments to impose restrictions on the mobility of citizens’ to minimize the risk of infection. This has led researchers and several news media to speculate about whether working remotely will be the new normal. Our project would try to identify whether there has been a significant increase in the number of job openings where the option of working remotely is available. If we do find such an increase, we could then look at these vacancies at a sector/branch level and this measure could be interpreted as a proxy of the level of flexibility each branch has in terms of working remotely, one could for example expect that jobs within the service and hospitality branch cannot be fulfilled remotely.

Most surveys and studies have focused on the employee side of working remotely, by looking at job postings we can have a glimpse into employer’s perception on the possibility of employees working remotely.

**Exam project**

At the end of the course your group must hand in a independent exam project.

The content of the exam project is something that you choose. You and your group must find a subject, data, choose methods etc.

**Grading**

The grade for this course is exclusively determined by the project handed in. The project will be judged on a number of dimensions, these include:

* how the data was obtained (setting up new data collection);
* how the data was processed;
* how machine learning methods are applied and which methods are used;
* how results are explained (writing, figures, tables with model output etc.);
* the research question and its originality as well as how it is answered.

Some advice about the grading. It is essential that spend time on motivating your project and conveying your results. In addition, it is important that you spend time on calibrating and validating the models you work with rather than using as many models as possible. We emphasize that using machine learning is NOT necessary to make a great project, many of the best projects gain insights from the data without modelling.

**Requirements for project**

The exam projects have a number of requirements that must be met, these are: requirement

* Research question (you should discuss with TAs)
* Groups with two to four members
* Project formalia
  + Project must consist of a report (.pdf file) and a documentation as Jupyter Notebook (.ipynb file).
  + The report should be written like a brief research article (short literature review, references to methods, results etc.). The report is limited to the following maximum number of pages (normalsider).
    - 2 members, 12 pages;
    - 3 members, 15 pages;
    - 4 members, 18 pages.
    - Note that 1 page (normalside) corresponds 2,400 and does not count figures, abstract, list of reference, frontpage, appendix.
    - The report should contain your exam numbers and possibly your names (optional). The exam numbers (or names) MUST show who contributed with writing which parts of the report. At most 20 pct. of the report can be written shared. If you fail to provide this the submission of your project may get rejected!
  + Grading will be based on the report but process but data collection, computations etc. should be well documented in the supporting Jupyter Notebook.

**Possible data sources**

Students in previous years of Social Data Science have used a large variety of data sources including:

* news on [DR](https://dr.dk/) (Danish Broadcasting Company) and the Danish newspaper Information
* price of cars for sales on [bilbasen](https://bilbasen.dk/)
* analyzing linguistic content on Twitter
* Airbnb pricing in Copenhagen
* Prediction of bitcoin prices from Reddit data.

If you are interested in working with one or more of these datasets or see the projects by the students who made them please contact us and we will put you in touch.

subject,

data,

methods

The report should be written like a brief research article (short literature review, references to methods, results etc.). The report is limited to the following maximum number of pages (normalsider).