**OPIM 5894 Survival Analysis Project**

**BACKGROUND**

1. Last year, you met Larry Hansen, COO of a pharmaceutical company, FermaLogis, during a networking event. You told him about your start-up company’s background in analytics and, especially, the survival analysis projects you have been involved and how survival analysis can make an impact on business. He told you that although he didn’t have a thorough knowledge of Survival Analysis, he started to hear the term more and more often in the recent years. He looked very interested. You exchanged contact information and parted. After seven months, you received an e-mail from Larry. The message is below:

*“Dear ......,*

*I hope you are doing well and life is good in CT. I know it is summer time, and you have your own obligations, but I need a small favor from you. A few weeks after we chat at the meeting, I encountered an essay on Survival Analysis on LinkedIn by an esteemed colleague of mine. He wrote about many uses of survival analysis in many industries. Encouraged by the essay, I went through the theory part of it. It turns out that we have a problem which can be solved by Survival Analysis. Yet, neither my team’s mathematics skills nor data science experience is competent enough to undertake such a technical project. I was thinking about whether our company can get some help from your capable team.*

*As I mentioned in my speech at the event, employees are very important for the FermaLogis operations as the main resource and input. So, we care about our employees so much. However, there is this issue that our employees are leaving the company more then we desire. Let me elaborate:*

*a. Young employees: We are investing a lot in the new employees in terms of professional training and salary without return for the first three years. Only after three years, we are starting to benefit from them. But, at that time, they become more competent professionally and seek new career opportunities with higher salaries and leave the company. This is a big problem because of two reasons. First, we can’t harvest what we invested and second, a talent gap hinders our operations.*

*b. Experienced employees: As I told you in the meeting, our company has an Executive Training Program. It is a very intense and useful program. Every year, our experienced employees (with the company for 5 years or more) are sent to a training center for two weeks and trained there by professors and consultants from top universities and firms. They learn valuable skills and become more competent each year. As I mentioned above, we have enjoyed the benefits of the training program a lot. But, on the other hand, they become the hot commodity for the rival companies and, recently rival companies are recruiting our well-trained employees with high transfer fees and higher salaries. We definitely don’t want this to happen for many reasons you can understand.*

*The reason why I am contacting you is that we want you to find out who leaves the company and why they are leaving. If we can understand the underlying reasons, we believe we can take the necessary measures to prevent our employees from leaving. I think that would be a very good application field for survival analytics. I am sending you the employee data (“FermaLogis\_Event\_Type.csv”).*

*One other thing is we want to relate our past bonus pays with turnover so that we can use bonuses more effectively to prevent turnover. We think our bonus payments don’t play an important part in employee turnover, but you can never know. I provide a data dictionary (“data\_dictionary.txt”) together with the dataset. Our database administrators prepared this dataset specifically for this project. There are still many issues with the data. Unfortunately, the only person who could answer your questions, our Head of Data Administration, has left the company three days ago. So, we have no further information about the dataset than the data dictionary. When you do your analysis, if you have any questions, please assume the most likely case, and proceed with the analysis under that assumption. Also, include any assumption you make throughout the project in your report.*

*If you are interested, please take a look at the data and let me know what we can do about this.*

*Thanks,*

*Larry Hansen*

*Chief Operating Officer at FermaLogis*

”

1. You take a look at the data and see that this data is sufficient to conduct a study on employee attrition. Following a trip to FermaLogis HQ, you agree on the project details, fees, and all the other payments. Your company, ......, writes and delivers a report on the following subjects.

Who is leaving the company?

Why are they leaving?

When is the biggest danger for employees to leave?

Is there any difference in attrition between different employee groups or categories?

Any other findings of the company.

Can I combine different event types together? Or do all need to be handled separately?

What attributes increase/decrease the hazard rates for certain event types?

Does bonus affect employee turnover? If yes, how?

Are there any variables which affect hazards non-proportionally?

**Any other question you think makes the report more complete.**

1. The project deadline is mentioned in the syllabus. Late submissions will receive a penalty as mentioned in the syllabus.
2. Grading will be based on your report. You must answer the general business needs of the company. The above questions on item 2 are only examples and they should be enriched.
3. As stated in Larry’s letter, FermaLogis cannot answer your questions about the dataset. You have to make assumptions, and shape your analysis based on those assumptions.