



# *An End-to-End Data Science Project*

---

**Deena Gergis**  
**Lead Data Scientist @ Bayer**



# Workshop overview:

## Session 1 Preparation 10.04.2022

*Start with the business problem, find data source, preprocess data, set up team process and tech*

## Session 2 Analytics 17.04.2022

*Analyze and understand your data. Gain insights and prepare for the predictive modeling*

## Session 3 Machine learning 08.05.2022

*Build and evaluate prediction model(s), use Mlflow to keep track of the various experiments*

## Session 4 Production 15.05.2022

*Create prediction functions and production class, develop an API, create a dashboard that the user will access and call the API*

## What you will do:

- **Form a team of 3 members**
- **During the sessions:** You will get tasks to be done
- **After the sessions:**
  - You will complete the whole covered phases
  - Dig deeper into the various technologies discussed



# *Refresher: Modelling Training*

- 1. Clean your data*
- 2. Decide on your modelling strategy*
- 3. Decide about the evaluation metric*
- 4. Train a baseline model*
- 5. Train more sophisticated models*
- 6. Decide which model will be used*



***But how we will use this model?***

***1. Predictive analytics:***

**Prediction of matching jobs given a skill set**



*Are notebooks sufficient for  
production use?*

**No**



*Refactor the code to  
proper scripts*



# ***But how we will use this model?***

## ***1. Predictive analytics:***

**Prediction of matching jobs given a skill set**

## ***2. Prescriptive analytics:***

**Simulate and recommend new skills**



# *Steps so far?*

1. *Notebooks: Modeling code*
2. *Mlflow: Models*
3. *Notebooks: Production dummies*
4. *Scripts: Refactoring #3*
5. *Then?*



# *Steps ahead*

*6. Scripts: API*

*7. Development in Web App*

*8. Monitoring & Retraining*





# *Other production standards?*

In addition to the Web Apps / Dashboard

- \* Scheduled processing
- \* Integration into Business Software
  - e.g. SAP



# Assignment

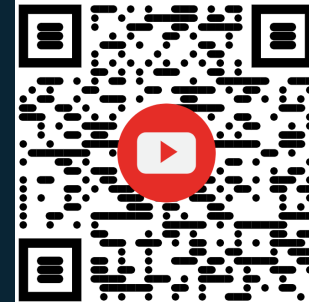
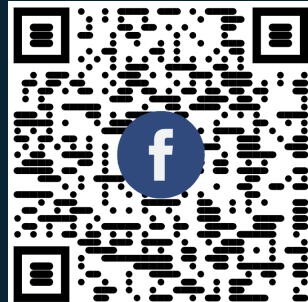
1. *Develop the prediction + simulation scripts*
2. *Build an API to expose those functionality*
3. *Build a web app as a product*



*Let's wrap up  
on Github*



# *And now it's your turn ... Questions?*



**LinkedIn:** [de.linkedin.com/in/deena-gergis](https://de.linkedin.com/in/deena-gergis)  
**YouTube:** [youtube.com/c/DeenaGergis](https://youtube.com/c/DeenaGergis)  
**Facebook:** [facebook.com/deena.gergis](https://facebook.com/deena.gergis)