ABSCHNITT 7

DIE HÄNDE SCHMUTZIG MACHEN

EINE PROGRAMMIERÜBUNG ZU ML IN PYTHON

Alina Herderich

FÜR DEN REST DES NACHMITTAGES

Wählt eine von drei Programmierübungen und fangt an, euch mit Machine Learning in Python vertraut zu machen. Ihr könnt in Kleingruppen (zwei bis drei Personen) zusammenarbeiten, um euch gegenseitig zu unterstützen.

ÜBUNG 1

Survival on the Titanic

ML Basics
die klassische Data Science Pipeline
Logistische Regression & Random Forest

ÜBUNG 2

Bags of Popcorn

Kaggle Tutorial
klassisches NLP
Word Embeddings (Word2Vec)
Sentiment Analysis

ÜBUNG 3

Hugging Transformers

Hugging Face Tutorial
Transformers
Transformers verwenden und fine-tunen

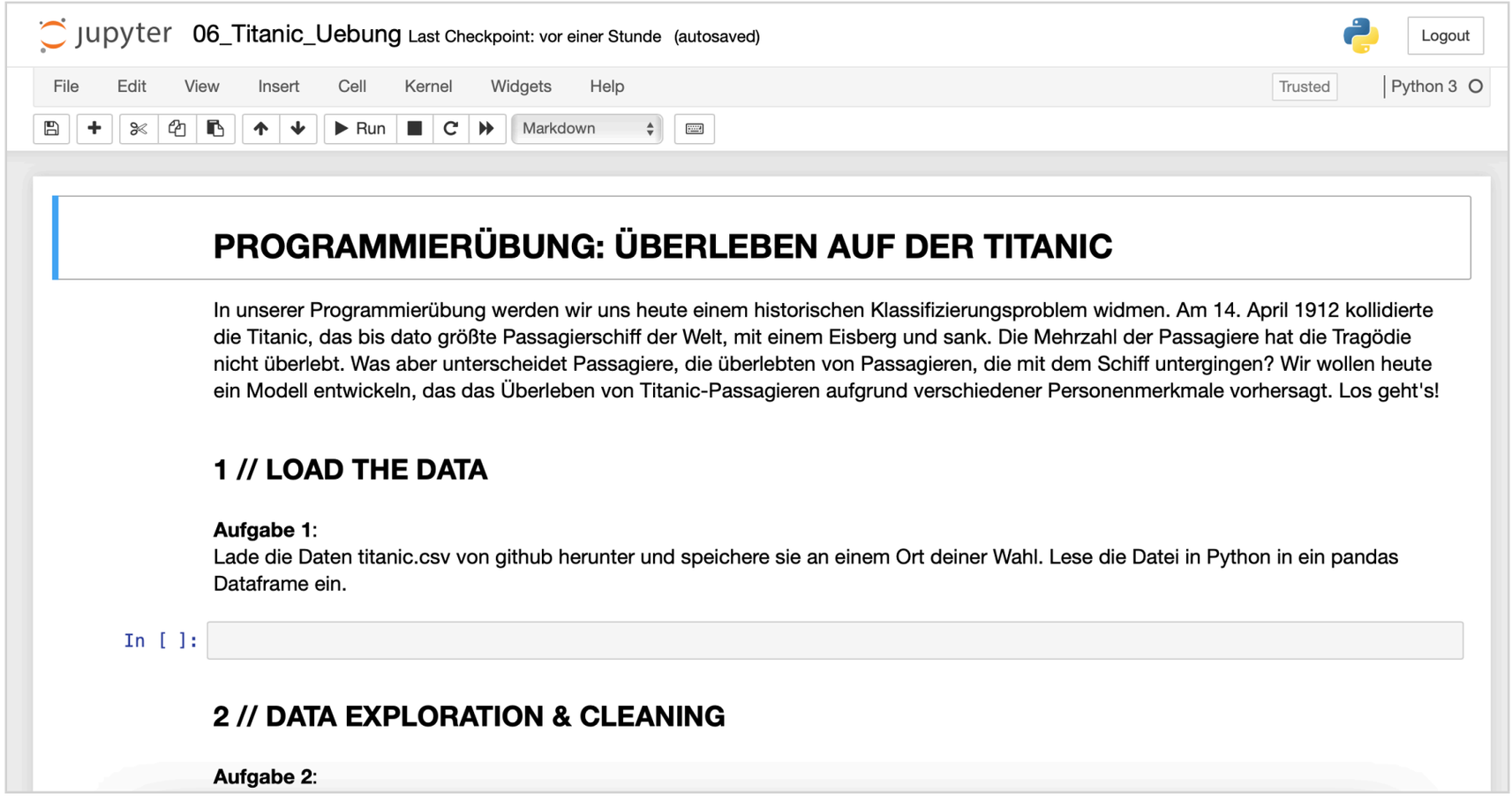
ÜBUNG 1:
SURVIVAL ON THE TITANIC

WELCHE MERKMALE BESTIMMEN DAS ÜBERLEBEN VON TITANIC PASSAGIEREN?

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

https://www.kaggle.com/c/titanic/data

WELCHE MERKMALE BESTIMMEN DAS ÜBERLEBEN VON TITANIC PASSAGIEREN?



WELCHE MERKMALE BESTIMMEN DAS ÜBERLEBEN VON TITANIC PASSAGIEREN?

Bestandteile der Übung:

explorative Datenanalyse & Plotten

fehlende Werte ersetzen

feature engineering

train-test split & cross validation

logistische Regression & Random Forest

Hyperparameter tunen

Modellevaluierung & Vergleich

WELCHE MERKMALE BESTIMMEN DAS ÜBERLEBEN VON TITANIC PASSAGIEREN?

Benötigte Packages:

pandas

matplotlib

seaborn

sklearn

ÜBUNG 2:
BAGS OF POPCORN

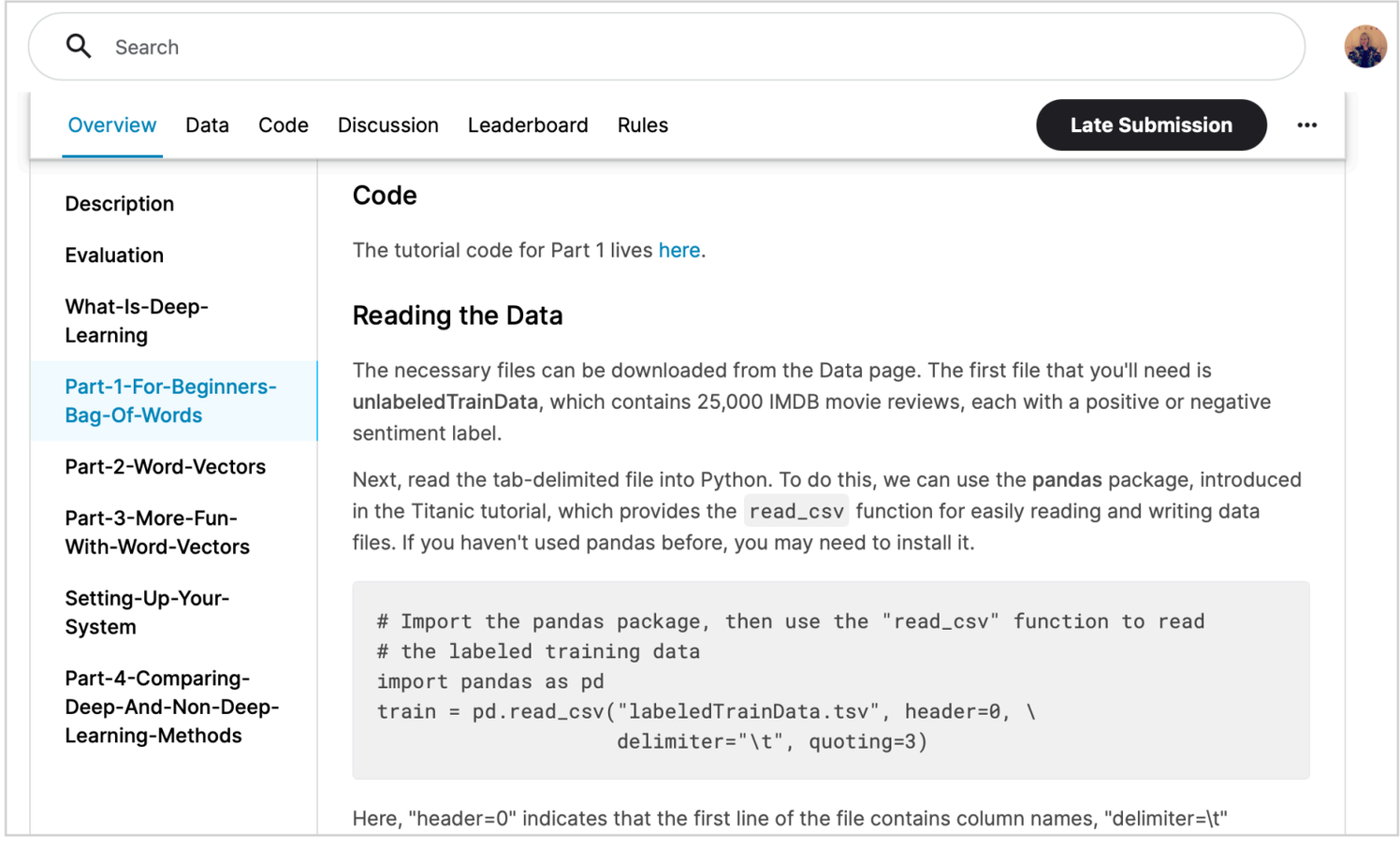
IST DIE VORLIEGENDE FILMBEWERTUNG POSITIV ODER NEGATIV?

Data fields

- **id** - Unique ID of each review
- **sentiment** - Sentiment of the review; 1 for positive reviews and 0 for negative reviews
- **review** - Text of the review

https://www.kaggle.com/c/word2vec-nlp-tutorial/data

IST DIE VORLIEGENDE FILMBEWERTUNG POSITIV ODER NEGATIV?



https://www.kaggle.com/c/word2vec-nlp-tutorial/overview/part-1-for-beginners-bag-of-words

IST DIE VORLIEGENDE FILMBEWERTUNG POSITIV ODER NEGATIV?

Bestandteile der Übung:

Datenreinigung (regular expressions, stop words, …)

tokenization & bag of words

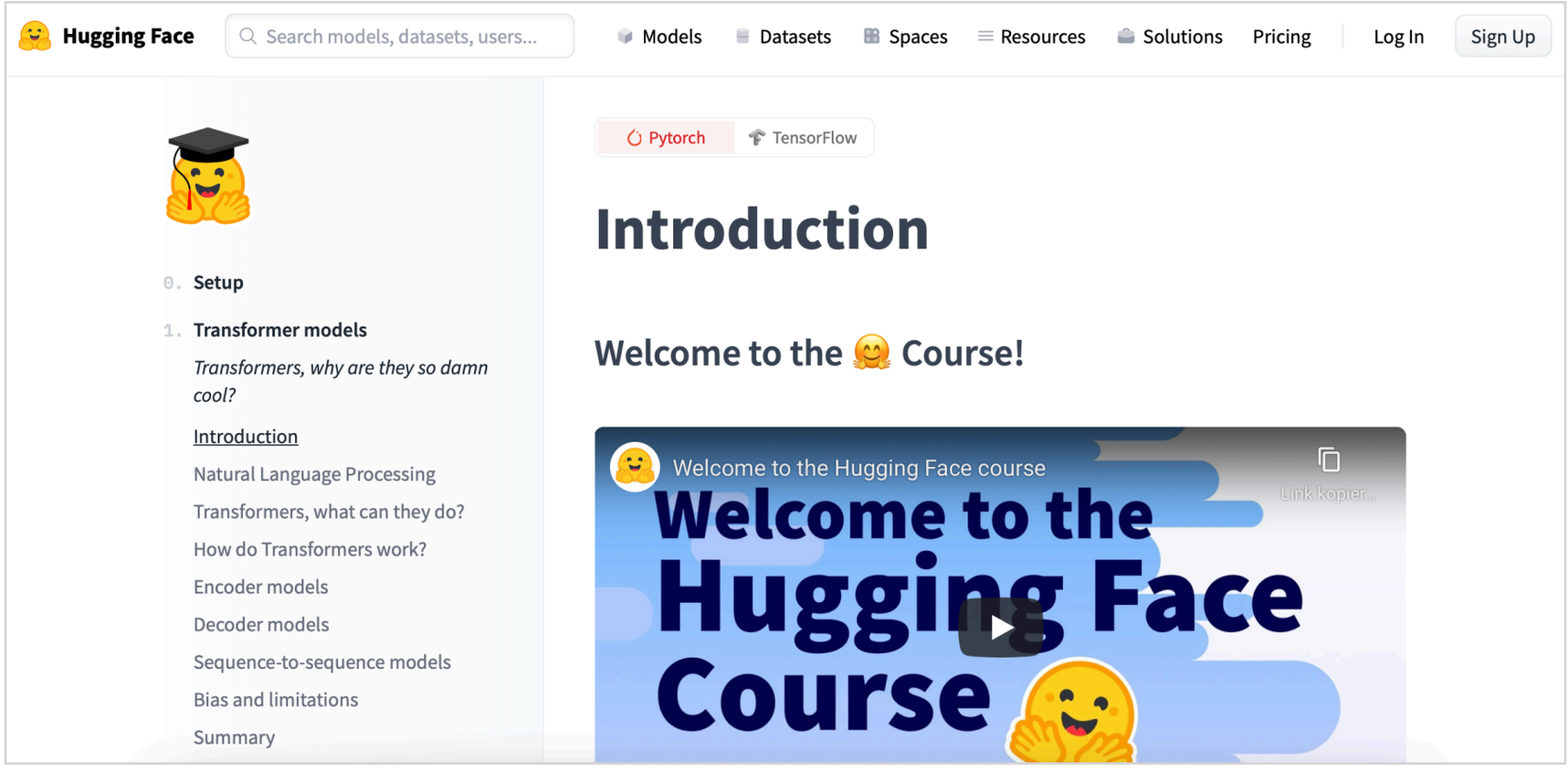
word embedding training (Word2Vec)

document representations (averaging, centroids)

Random Forest

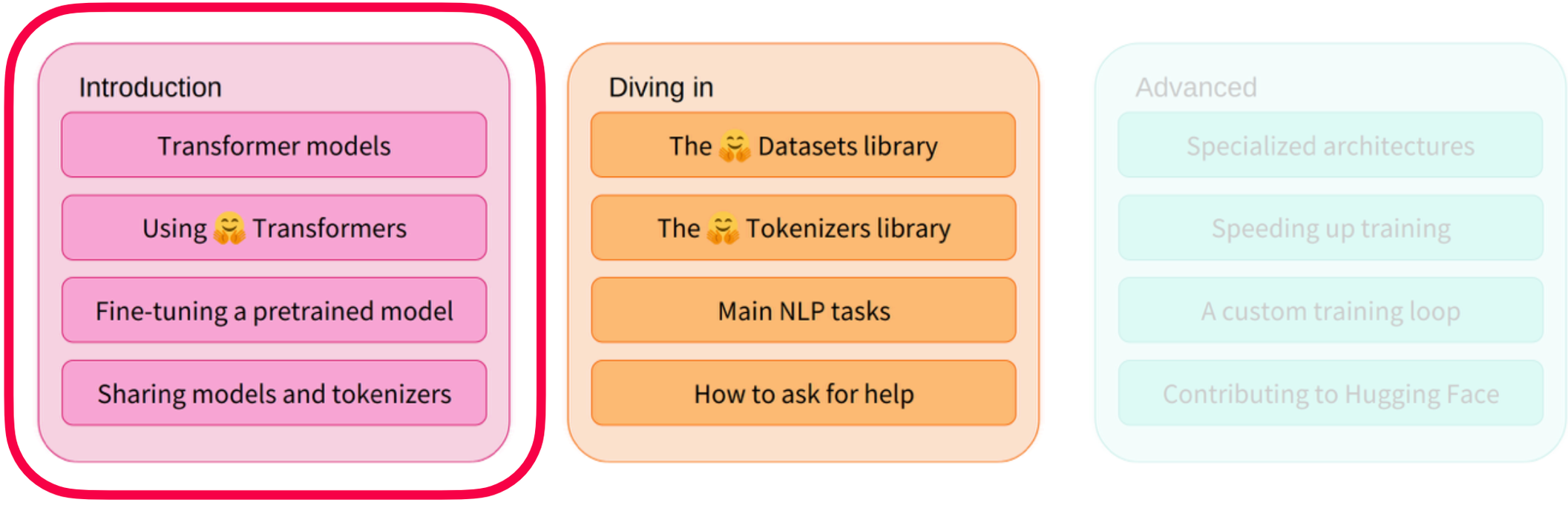
ÜBUNG 3:
HUGGING TRANSFORMERS

HUGGING FACE: DIE TRANSFORMER LIBRARY ERKUNDEN.



https://huggingface.co/course/chapter1/1

HUGGING FACE: DIE TRANSFORMER LIBRARY ERKUNDEN.



hier anfangen

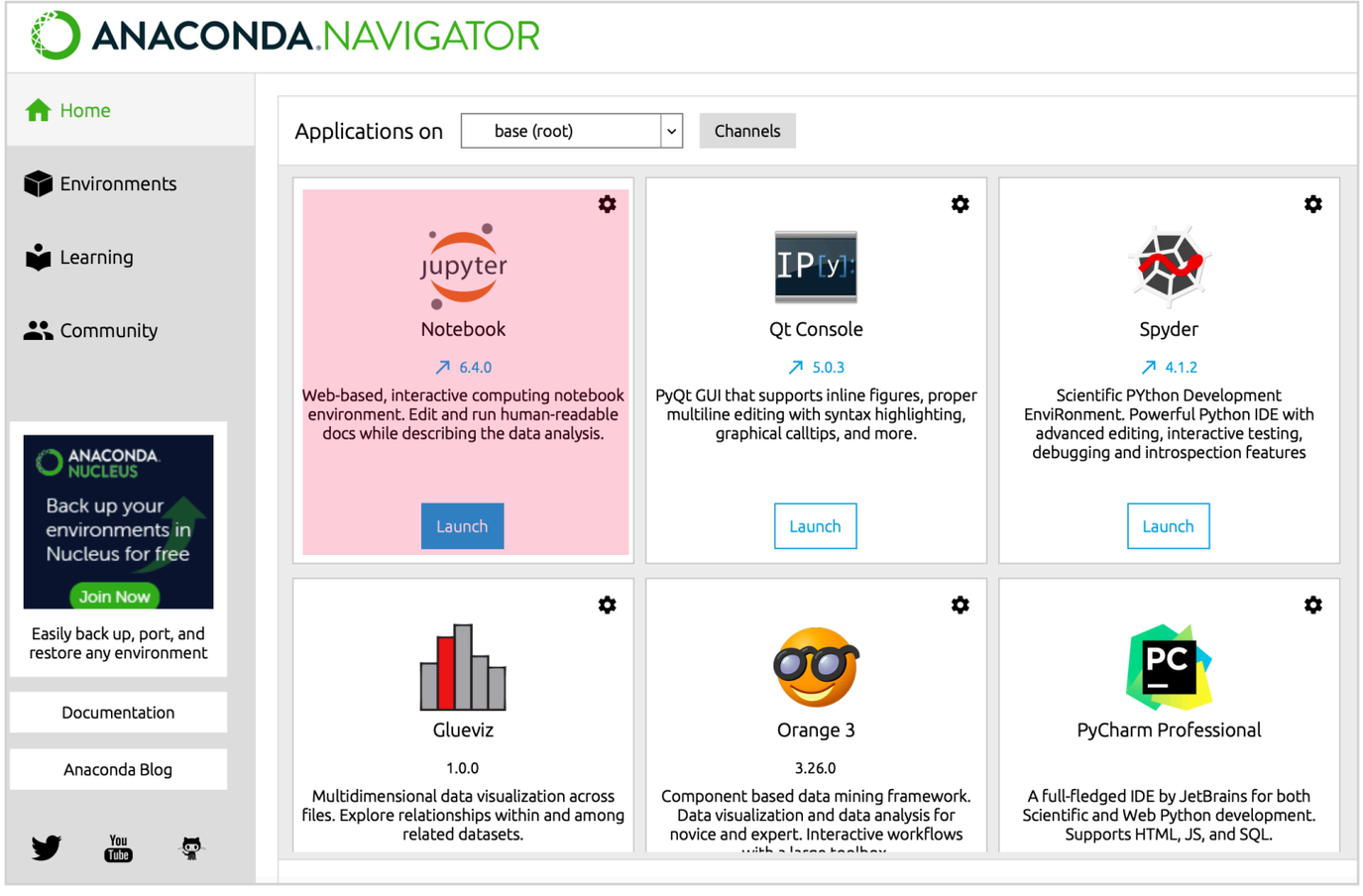
sehr wahrscheinlich nicht mehr für heute

noch nicht released

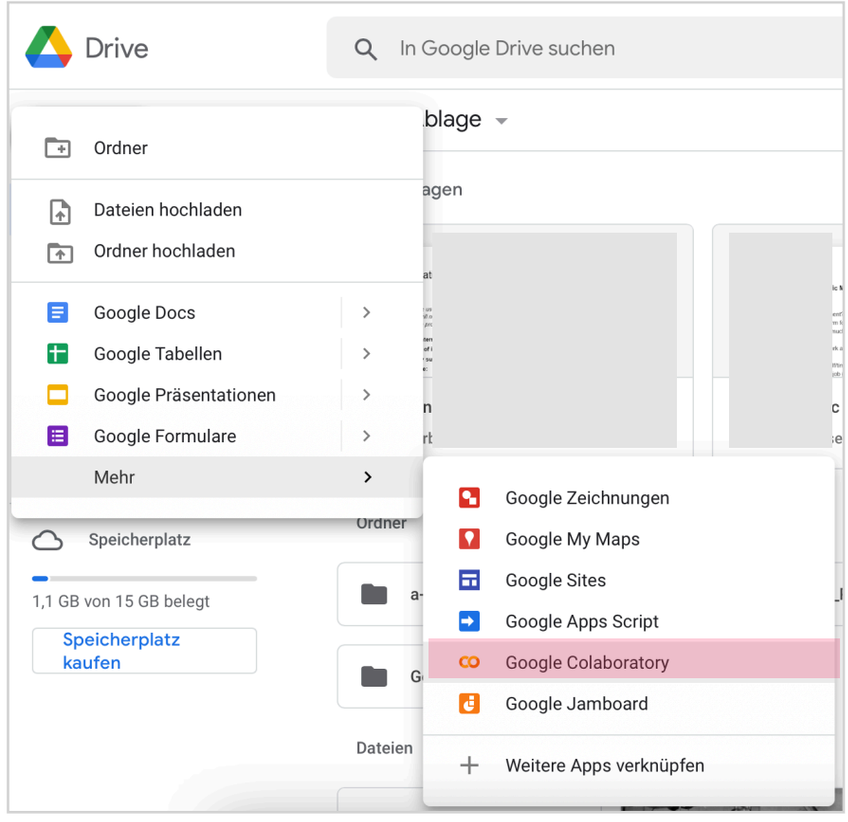
https://huggingface.co/course/chapter1/1

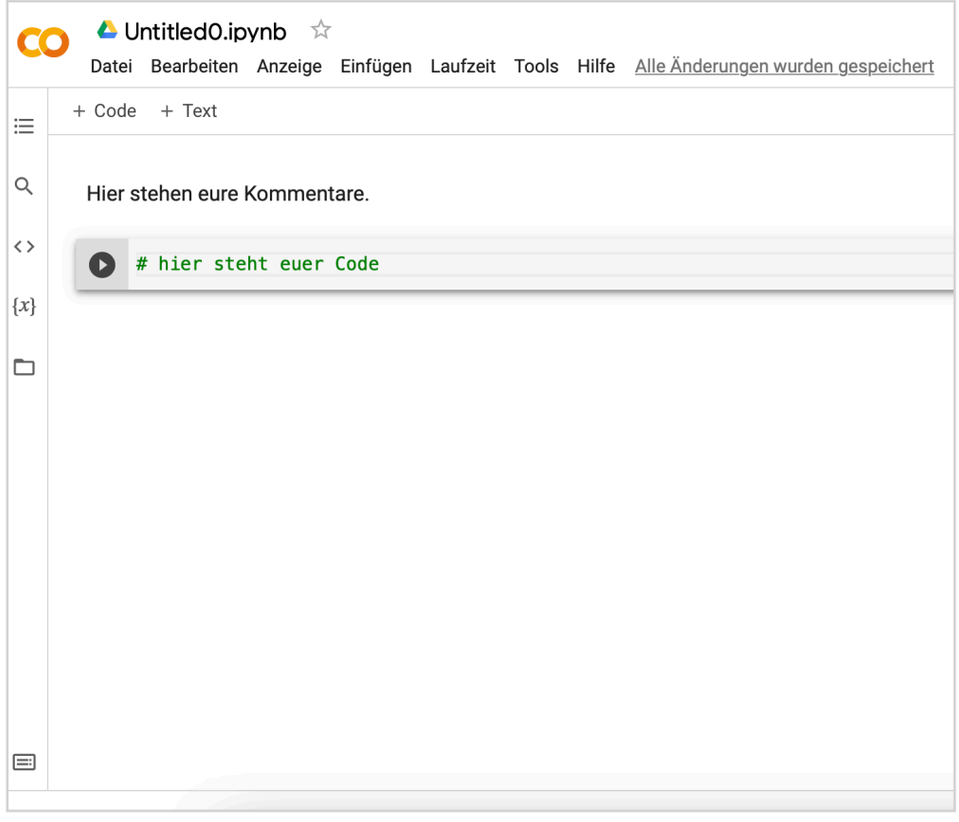
ICH HABE NOCH KEINE PROGRAMMIERUMGEBUNG ODER BIN MIR MIT DEREN UMGANG UNSICHER.

MÖGLICHE UMGEBUNGEN FÜR DIE ÜBUNG: ANACONDA & JUPYTER NOTEBOOK

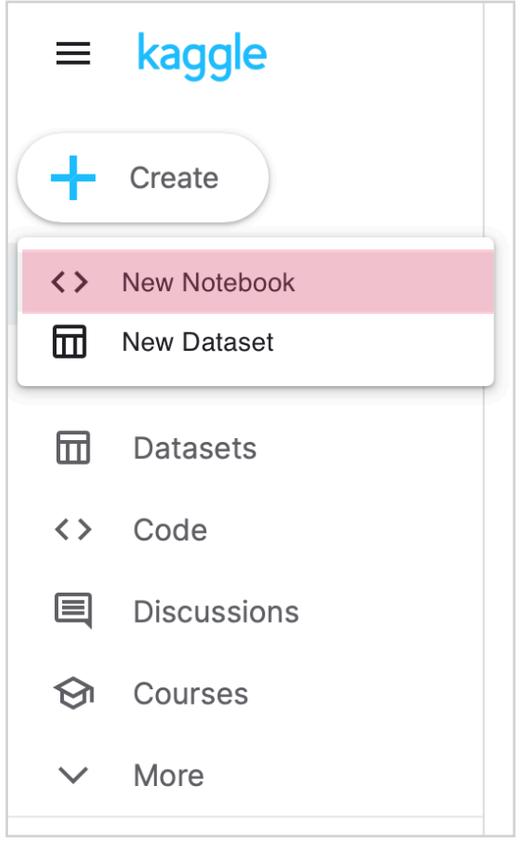


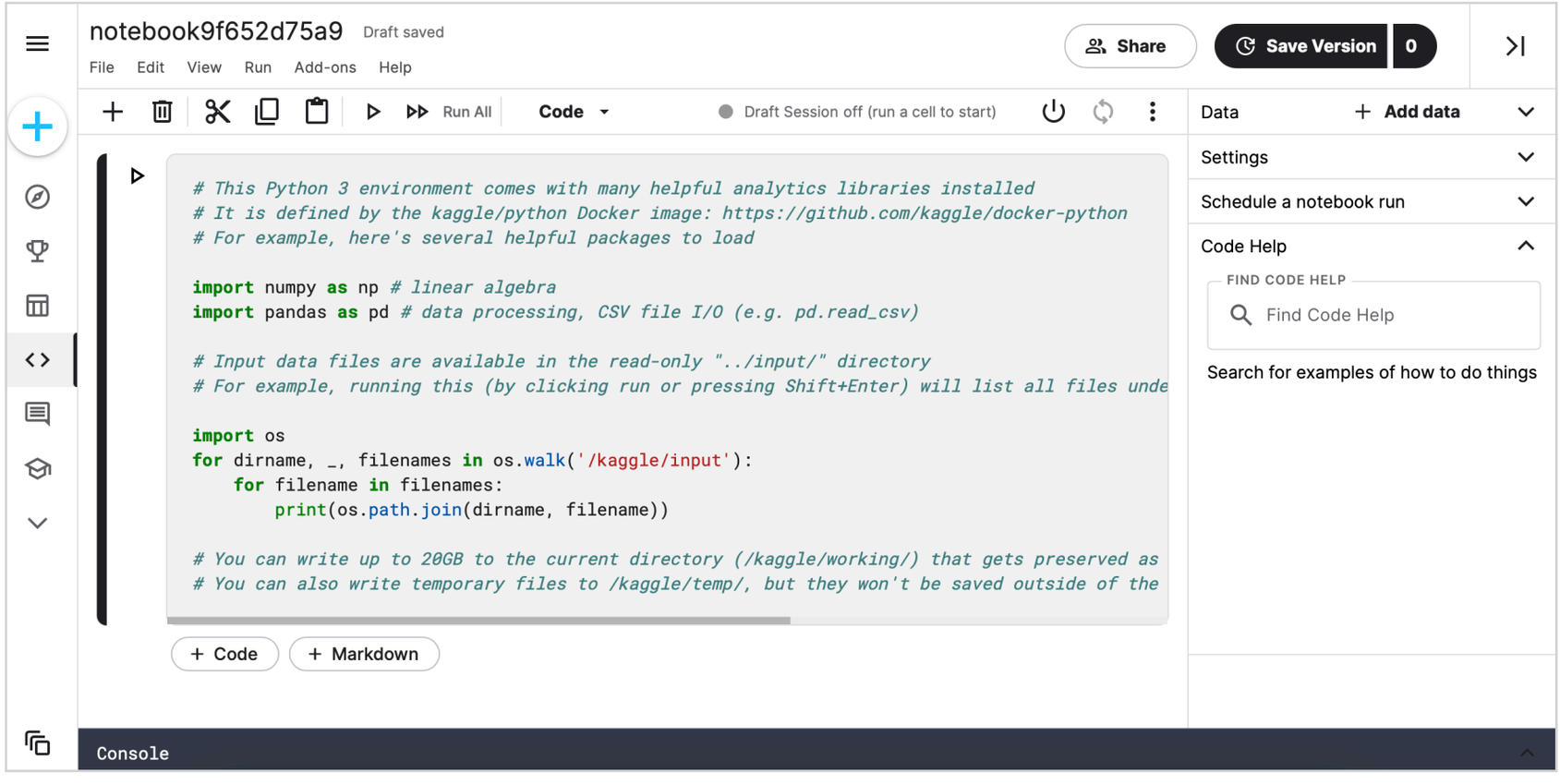
MÖGLICHE UMGEBUNGEN FÜR DIE ÜBUNG: GOOGLE COLAB





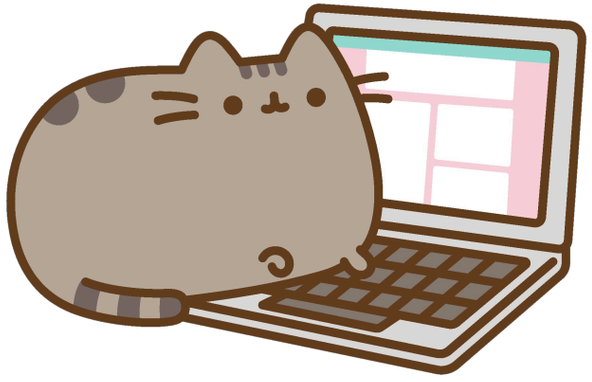
MÖGLICHE UMGEBUNGEN FÜR DIE ÜBUNG: KAGGLE





GUT, DANN NEHME ICH ÜBUNG …

PAUSE NACH BEDARF.



LOS GEHT’S!