



# Introducing PyCaret

July 17, 2020



# About me

- Data Analytics Leader by day and passionate Data Scientist by night
- Active open-source contributor
- Member of CPA, CMA, Canada and ACMA, UK, CGMA
- Lived and worked in Asia, Middle East, East Africa, North America
- Currently based in Toronto, Canada



<https://www.linkedin.com/in/profile-moez/>

**Medium**

[https://medium.com/@moez\\_62905/](https://medium.com/@moez_62905/)



<https://twitter.com/moezpycaretorg1>



[moez@pycaret.org](mailto:moez@pycaret.org)

# Important Links

- Official : <https://www.pycaret.org>
- PyCaret GitHub : <https://www.github.com/pycaret/pycaret>
- LinkedIn : <https://www.linkedin.com/company/pycaret>
- YouTube : [https://www.youtube.com/channel/UCxA1YTYJ9BEeo50lxyl\\_B3g](https://www.youtube.com/channel/UCxA1YTYJ9BEeo50lxyl_B3g)
- Medium : [https://medium.com/@moez\\_62905/](https://medium.com/@moez_62905/)
- Today's Presentation and Demo : <https://www.github.com/pycaret/pycaret-demo-at>

Follow hashtag #pycaret on LinkedIn and Twitter

# Agenda

- What is PyCaret?
- Demo using PyCaret 1.0x
- New Features in PyCaret 2.0x
- Feature Demo using `pycaret-nightly`
- Q&A

# What is PyCaret?

PyCaret is an open source, low-code machine learning library in Python that aims to reduce the cycle time from hypothesis to insights. PyCaret can be used for rapidly developing and deploying machine learning pipeline.



**EASY TO USE**



**PRODUCTIVITY TOOL**



**BUSINESS READY**

# Effect of using PyCaret

Fig 1.2 - Coding in PyCaret

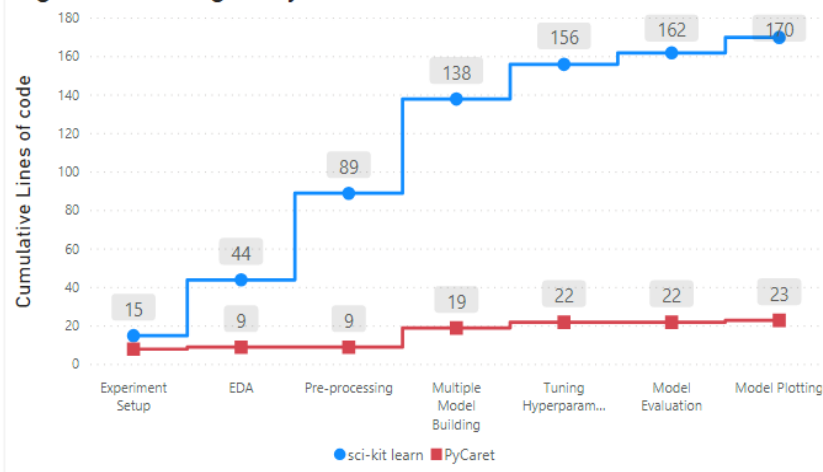
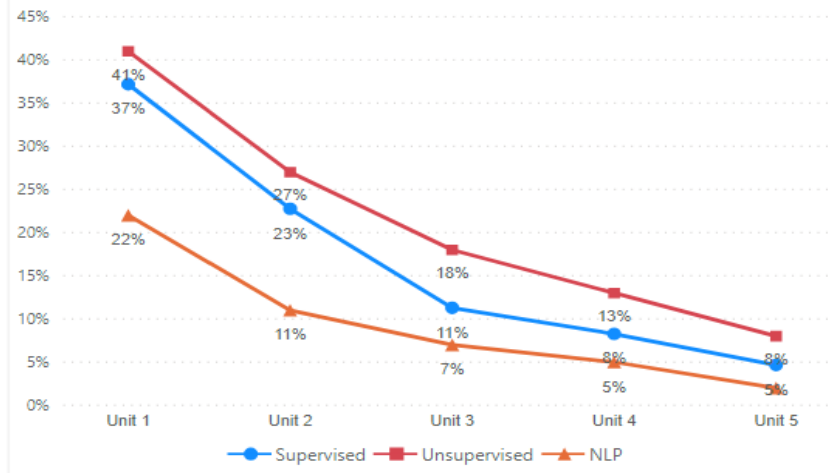


Fig 1.1 - Productivity gain using PyCaret



# Facts

- Started in summer of 2019
- Written in Python. Future versions may include wrappers in other langs.
- Project took approximately 1 year to complete.
- Self-funded so far.
- Supported by team of brilliant people. To contribute please visit <https://www.pycaret.org/contribute>

# Resources



Moez Ali in Towards Data Science  
May 27 · 10 min read



## Deploy Machine Learning Pipeline on Google Kubernetes Engine

A beginner's guide to containerize and...



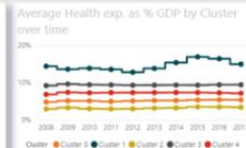
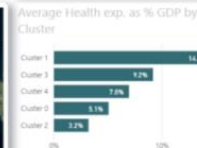
314

2 responses

<https://towardsdatascience.com/deploy-machine-learning-model-on-google-kubernetes-engine-94daac85108b>



Moez Ali in Towards Data Science  
May 20 · 8 min read



## How to implement Clustering in Power BI using PyCaret



83



<https://towardsdatascience.com/how-to-implement-clustering-in-power-bi-using-pycaret-4b5e34b1405b>



# Resources (cont.)



Moez Ali in Towards Data Science  
May 13 · 9 min read

Employee Credit Card Transactions 2014-2019

2/25/2019 12/20/2019



Anomalies over time



Divisions with high anomalies

DIV_NAME	Anomaly Score	# Anomalies
Academic Support	265	19
E Education Block Grants	36	1
EDUCATION BLOCK GRANTS	265	13
Educator Support	181	8

## Build your first Anomaly Detector in Power BI using PyCaret

A step-by-step tutorial for implementing...



379

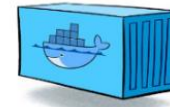
3 responses



<https://towardsdatascience.com/build-your-first-anomaly-detector-in-power-bi-using-pycaret-2b41b363244e>



Moez Ali in Towards Data Science  
May 7 · 12 min read



PyCARET



## Deploy Machine Learning Pipeline on cloud using Docker Container

A beginner's guide to deploy machine...



585

3 responses



<https://towardsdatascience.com/deploy-machine-learning-pipeline-on-cloud-using-docker-container-bec64458dc01>

# Resources (cont.)



Moez Ali in Towards Data Science  
May 1 · 11 min read



**PYCARET**

A beginners guide to train and deploy machine learning pipelines in Python



## Build and deploy your first machine learning web app

A beginner's guide to train and deploy machine...



532

2 responses



<https://towardsdatascience.com/build-and-deploy-your-first-machine-learning-web-app-e020db344a99>



Moez Ali in Towards Data Science  
Apr 23 · 15 min read



## Machine Learning in Power BI using PyCaret

A step-by-step tutorial for implementing machine learning...



768

9 responses



<https://towardsdatascience.com/machine-learning-in-power-bi-using-pycaret-34307f09394a>

# Resources (cont.) – Official Video Tutorials

- Binary Classification Video Tutorial

<https://www.youtube.com/watch?v=2xAgLKUN6Xs>

- Clustering in PyCaret Video Tutorial

<https://www.youtube.com/watch?v=2oxLDir7foQ>

- Anomaly Detection in PyCaret Video Tutorial

<https://www.youtube.com/watch?v=q0dxYDq1A40&t=2s>

- Topic Modeling in PyCaret Video Tutorial

<https://www.youtube.com/watch?v=G6ShuoM3T1M>

- Association Rule Mining in PyCaret Video Tutorial

<https://www.youtube.com/watch?v=XYAGwts5qGw>

# Resources (cont.)

- Machine Learning in SQL by **Umar Farooque**

<https://towardsdatascience.com/machine-learning-in-sql-using-pycaret-87aff377d90c>

- PyCaret's integration with Tableau by **Andrew Cowan-Nagora**

<https://towardsdatascience.com/machine-learning-in-tableau-with-pycaret-166ffac9b22e>

- NLP Classification using PyCaret by **Prateek Baghel**

<https://towardsdatascience.com/predicting-crashes-in-gold-prices-using-machine-learning-5769f548496>

- Predict Gold Price Returns using PyCaret by **Riazuddin Mohammad**

<https://towardsdatascience.com/machine-learning-to-predict-gold-price-returns-4bdb0506b132>

- Predict Crashes in Gold Price using PyCaret by **Riazuddin Mohammad**

<https://towardsdatascience.com/predicting-crashes-in-gold-prices-using-machine-learning-5769f548496>

# Resources (cont.)

- Binary Classification Notebook by **Murali Tedla**

<https://colab.research.google.com/drive/1W6ZYw5oAN7V85utFkXkCQq3iBWjqP1WB>

- Wine Quality Prediction by **Abhinav Arora**

<https://colab.research.google.com/drive/1uFNmhgM9ghVrtqN3MXaofxDILePb3yli#scrollTo=sYsHyY8uH6tb>

- Build your machine learning models by **Lakshay Arora**

<https://www.analyticsvidhya.com/blog/2020/05/pycaret-machine-learning-model-seconds/>

- PyCaret's Integration with Power BI (1 Hour Video Tutorial)

<https://www.youtube.com/watch?v=ZeQVKEWo2x0>

- Anomaly Detection using PyCaret by **Krish Naik** (Video Tutorial)

<https://www.youtube.com/watch?v=ZeQVKEWo2x0>

# Demo PyCaret 1.0x



1. Clone repository from <https://github.com/pycaret/pycaret-demo-at>
2. Install PyCaret by running "pip install pycaret" in Notebook or in Anaconda prompt
3. If you are using MAC OS, you might have to install LightGBM separately. Follow instructions on this link: <https://www.pycaret.org/install>

1. Links to demo notebook at <https://www.pycaret.org/demo>
2. You can also execute the notebook in playground mode by installing pycaret on colab by using "!pip install pycaret".
3. You must have a google account to use Google colab.

# New Features in PyCaret 2.0x



- **Modular Automation**

- Behavior of `compare_models`
- Behavior of `tune_models`
- Customize parameters
- NEW FUNCTION: `automl`

- **MLFlow logging back-end**

- NEW PARAMETER: `log_experiment`
- NEW FUNCTION: `get_logs`
- System logs automatically created

- **Command Line Support**

- Support outside Notebook environment
- Execute as python scripts
- Remote execution (for e.g. Kaggle Kernels)

- **Preprocessing**

- NEW PARAMETERS:
  - `fix_imbalance`
  - `data_split_shuffle`
  - `folds_shuffle`
  - `remove_perfect_collinearity`

- **Utility functions**

- `pull function`
- `models`
- `get_logs`
- `get_config`
- `set_config`

# Thank you

Questions?