

Data Science and Cognos Analytics

November 2019

BI and Data Science can leverage each other's skills to add value to data insights

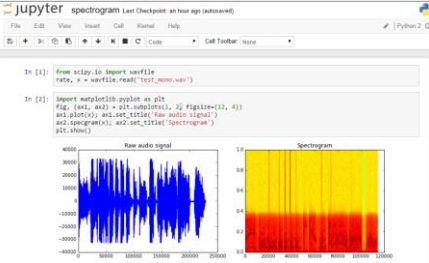
Data scientists often work in Silos and are not integrated with the BI ecosystem

- Decentralized
- Data access issues
- Unable share and collaborate widely

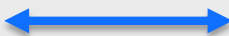
Enabling data scientists within the BI ecosystem has significant value

- Data scientists can build functions against governed and centralized data (Single version of the truth)
- Report authors can build reports shaped and transformed (ETL) by Python scripts
- Report authors can embed functions, custom visualizations etc. created by data scientists
- Data scientists can share their assets with other BI users with the right security, permissions and governance
- Empower all BI users with advanced predictive analytics

Jupyter notebook (data scientists)



Cognos Environment (BI Users)



Data

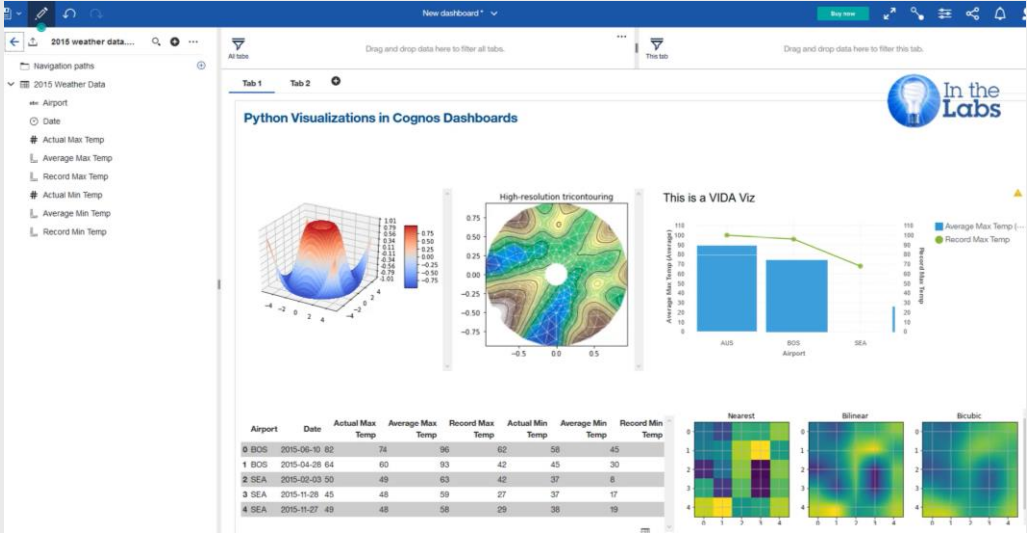


Why do we need to install our own Jupyter server as opposed to letting users connect to their existing Python environments?

- Sandboxing – we want to make sure that each user executing Python is sandboxed to their own environment. With something like Python any author could write malicious python code to harm or reformat a server. By giving each user his/her own environment we limit this malicious code to their environment only. The execution of any instance of the notebook is sandboxed for security
- Scalability – One single server won't get overloaded. We can scale out with new pods to accommodate traffic.
- Standardized python environment that allows all users to work on a production parity environment

Data Science and Cognos

- Notebooks as an artifact in CA
- Integrated with CA data (modules, data sets etc.)
- Publish shaped data to the Cognos content store
- Notebook integration to dashboards Upload existing notebooks to CA
- Schedule notebooks
- Scalable, secure and multi-user infrastructure



IBM