

How Fit is our Data ?

Team EC_03

Rajshekhar Mukherjee(Id: PRT 269)

Arslan Ahmed(Id: PRT 267)

Yanpeng Gao(Id: PRT 112)

Japhet Bukongo(Id: PRT 163)

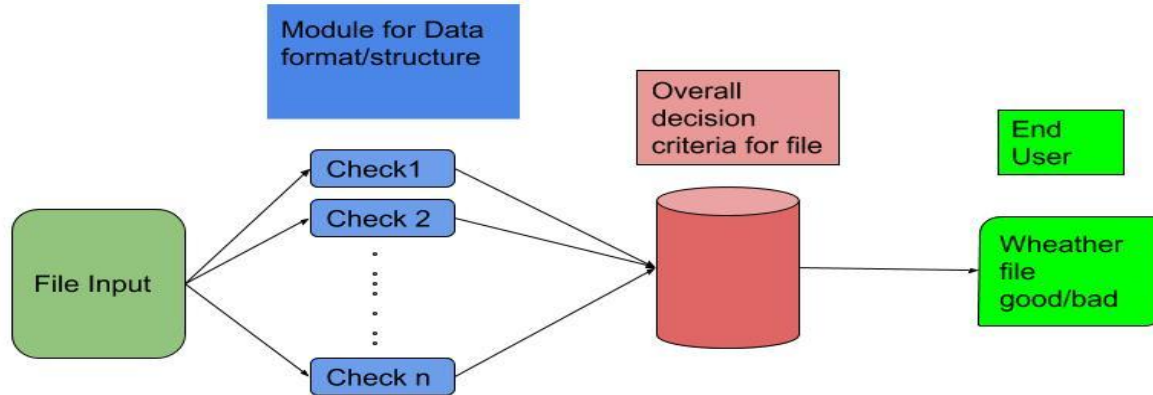
Business Case : Automating Data Fitness Assessment

- Inconsistent data quality and uncontrolled data flow pathways
- **FAIR** (Findable, Accessible, Interoperable, and Reusable) to optimize time spent by analysts and Data Scientists
- Develop an overall pass/fail dataset rating based on assessment results to conclude if the dataset is reuseable or not

User Story



Data Flow for Implementation



Technologies used

- Used Python for data fitness checking
- On Jupyter IPython notebooks (Google Colaboratory)

A decorative graphic on the left side of the slide, consisting of overlapping green and teal geometric shapes that form a stylized arrow pointing to the right.

Demo Starts...

- We use IPython notebook for demo