



1. Extract CSV files and store in data lake
2. Create databrick and establish mountpoint to clean and transform CSV data
3. Connect to APIs, read relevant data into dataframe, clean and transform data
4. Create databricks for Kafka producer and consumer, produce + consume Kafka messages to stream EPA API data
5. Construct SQL tables and read data from Azure blobs to SQL database
6. Set up data factory to trigger producer + consumer activity / create pipeline to update data in SQL
7. Create databrick and develop linear regression ML model using asthma + historic air quality data
8. Save ML model and dump into PowerBI / implement on EPA API data to get daily predictions
9. Create PowerBI dashboard incorporating above ML model + historic data visuals

