AIR QUALITY ANALYSIS

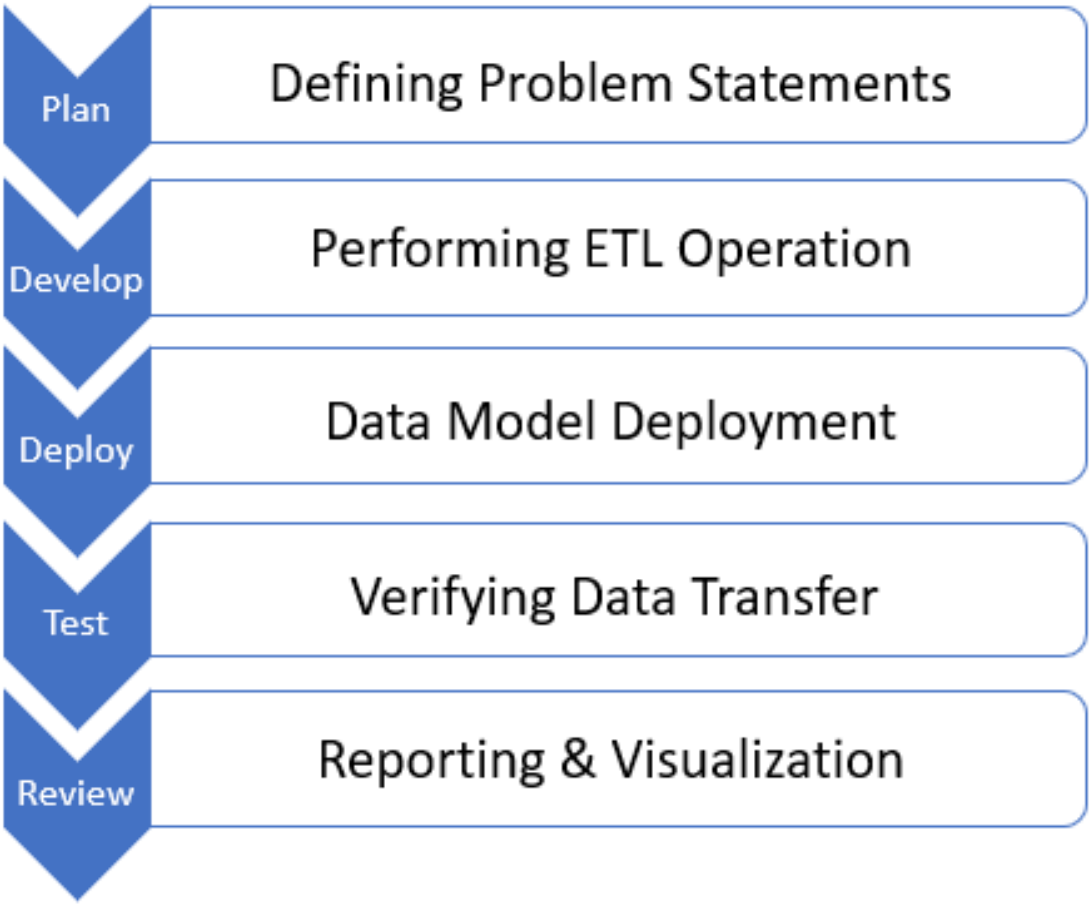
Submitted By:

Kshitiz Bansal

Problem Statements

1. Yearly & Monthly Air Quality Analysis
2. Detailed State & City Selective Air Quality Analysis
3. Impact Of Covid-19 On Air Quality
4. Observing Harmful Gases Content Variation
5. State Ranking On Yearly Basis
6. Analysing AQI Scale Over Period Of Years

Agile Methodology



Blueprint

Azure Data Factory

Extracting Data
From External Source
(HTTP)

Azure Databricks

Cleaning &
Transformation
Operations

Azure Analysis Services

Establishing Internal
Relationships & Data
Model Deployment

Powe BI

Visualization &
Reporting

Calculation Metrics

Metrics & Dimensions

- Geographical Location
- Time Dimension

Parameters

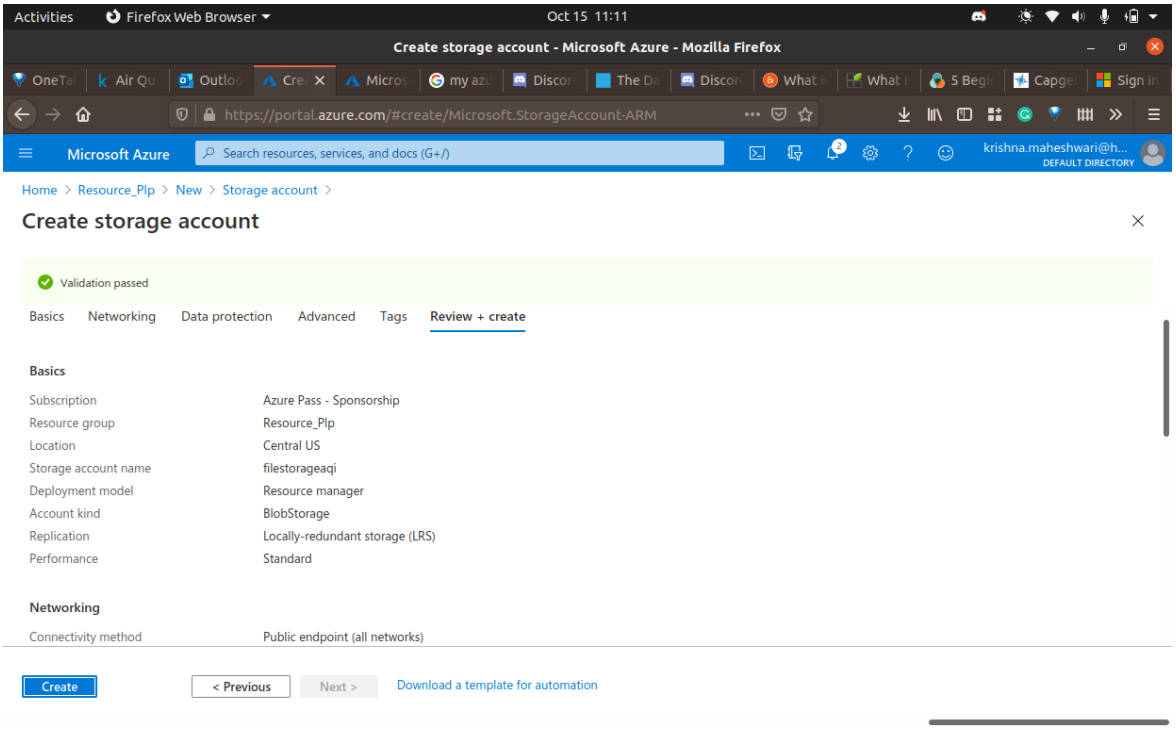
Data Parameters

- City
- State
- Datetime
- PM10
- PM2.5
- NO
- NO2
- NOx
- NH3
- CO
- SO2
- O3
- Benzene
- Toluene
- Xylene
- AQI
- AQI Bucket

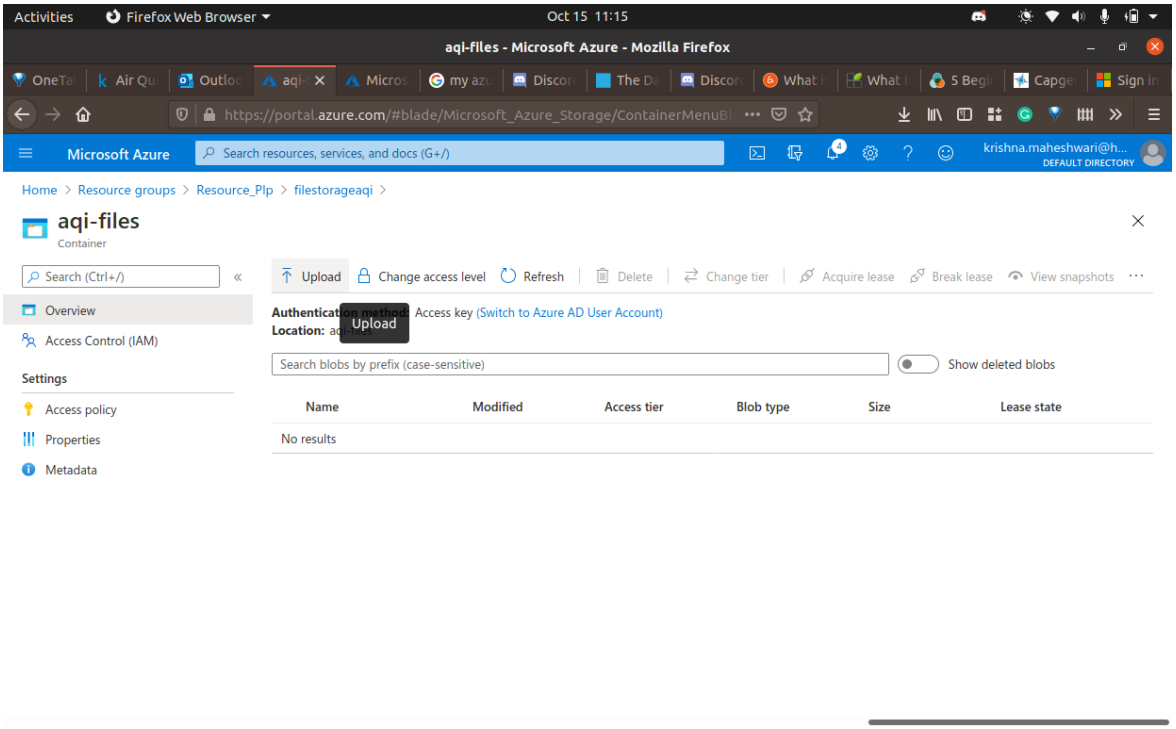
Extracting Data

1. Creating An Azure Data Factory Instance
2. Creating A Linked Service With HTTP Source (Kaggle)
3. Saving Extracted Data To Blob Storage
4. Creating A Databricks Notebook Activity
5. Mounting Blob Storage With Databricks Notebook
6. Reading Data Through Triggering Pipeline

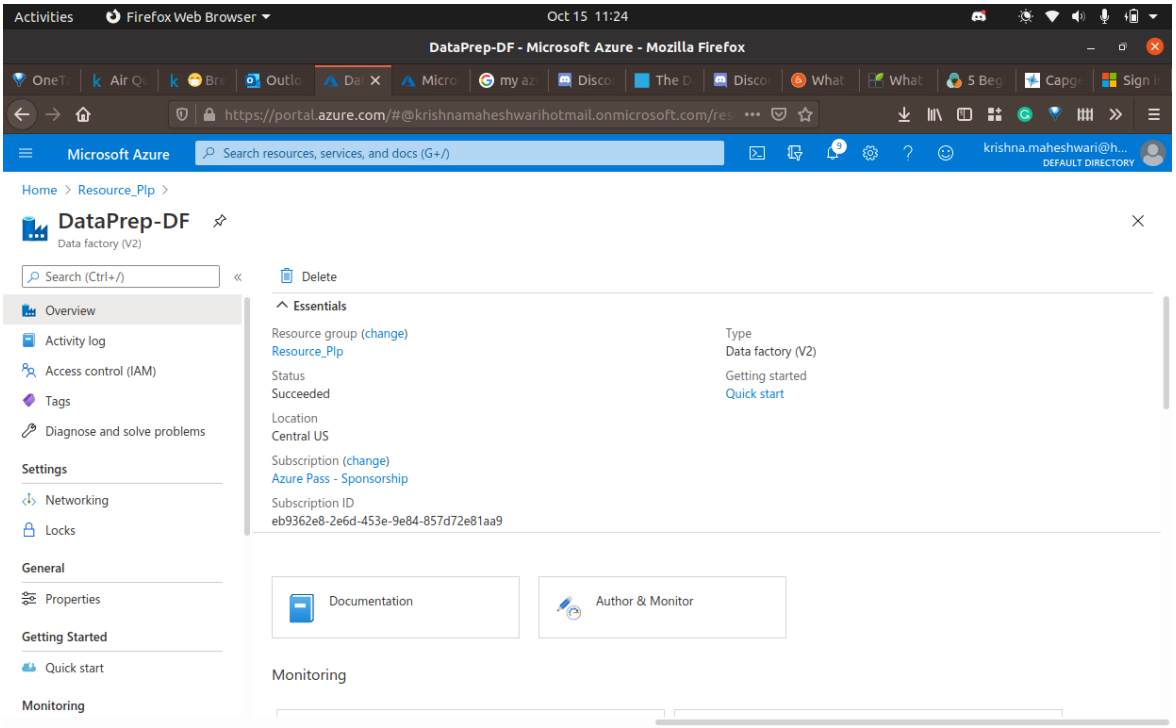
Creating Storage Account



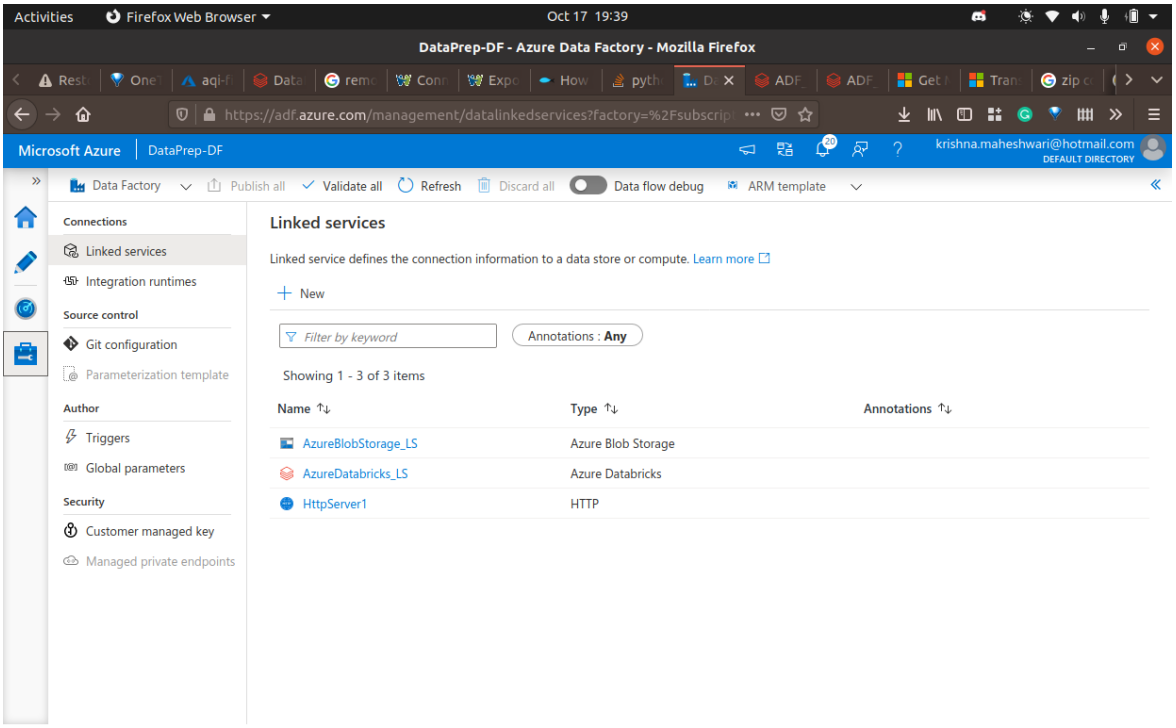
Creating Blob Container



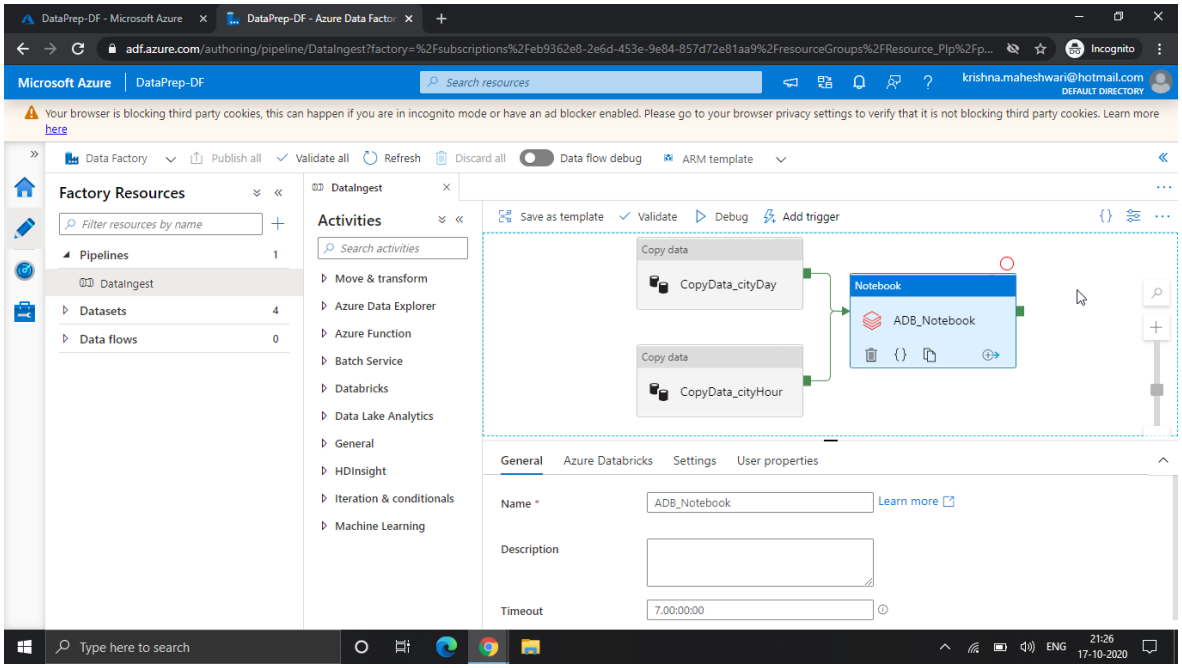
Creating Data Factory



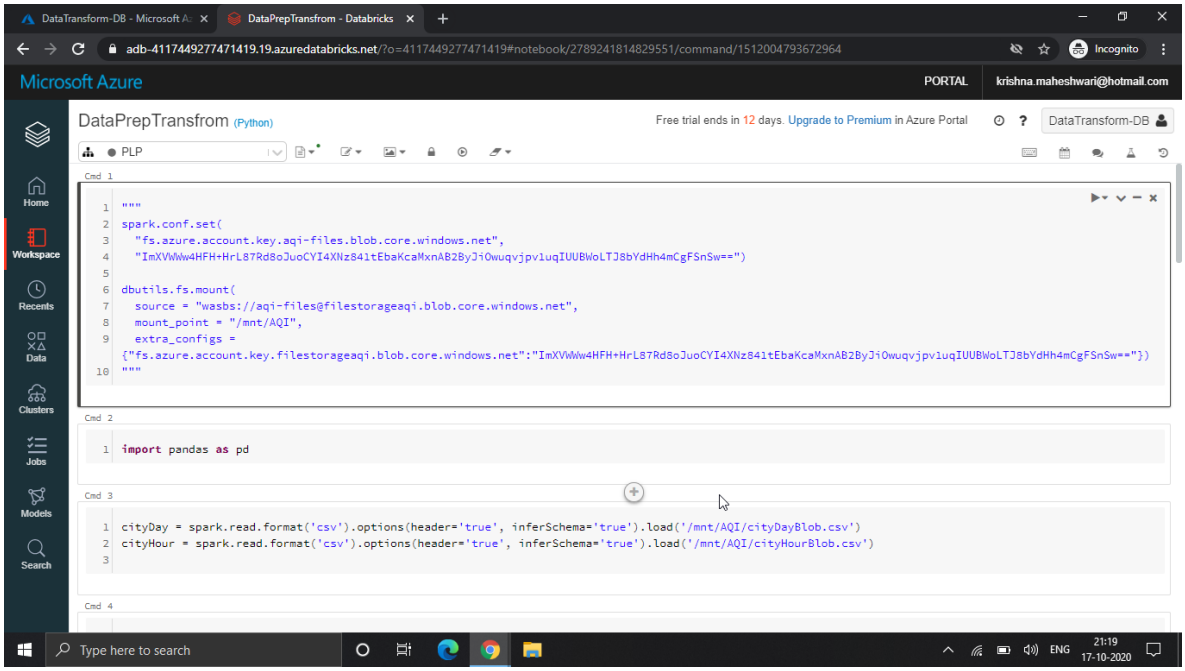
Creating HTTP Linked Service



Creating Databricks Notebook Activity



Mounting Blob Data To Databricks Notebook

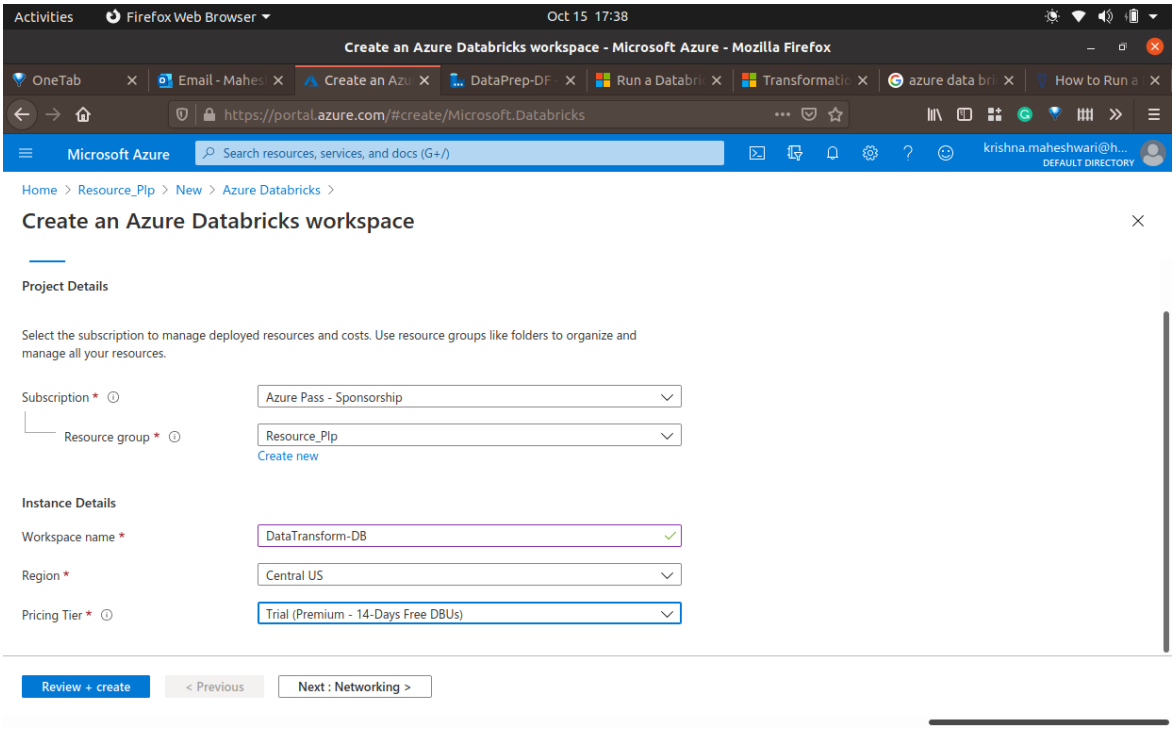


Data Transformation

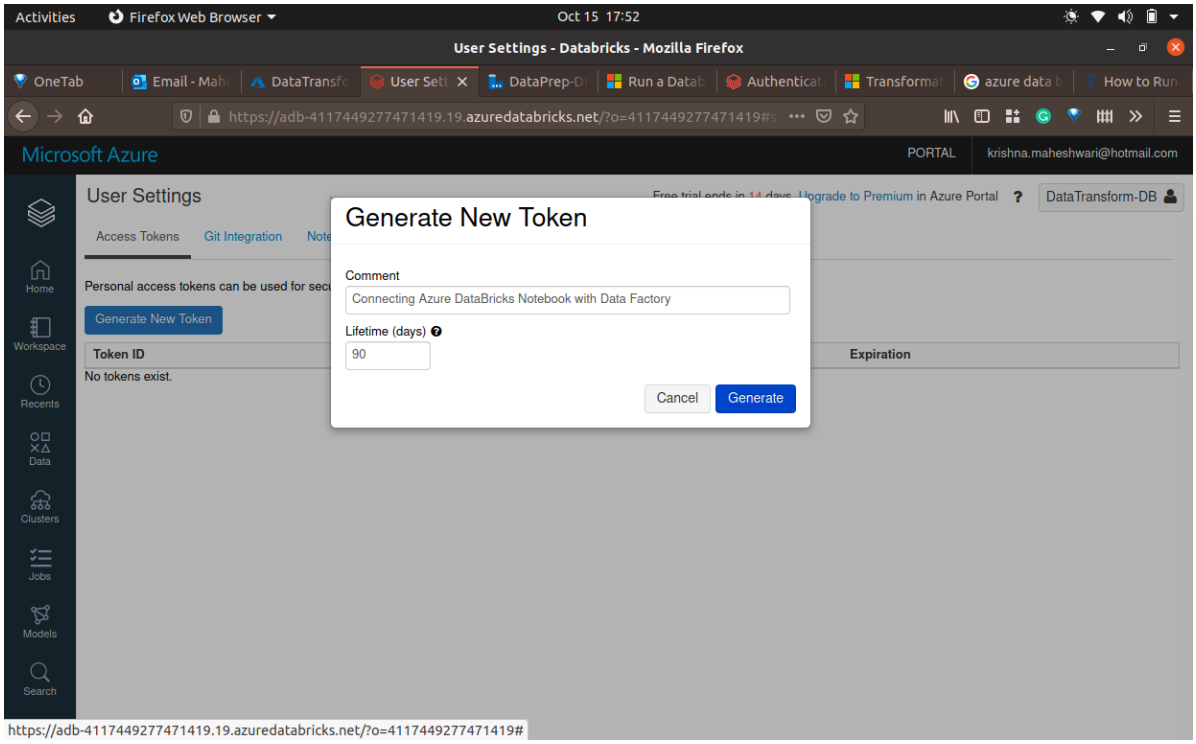
Language Used – Python 3

1. Detecting Schema For Each Dataframe
2. Using Linear Interpolation To Handle Missing Values
3. Creating Desired Calculated Columns
4. Mounting Transformed Data Back To Blob

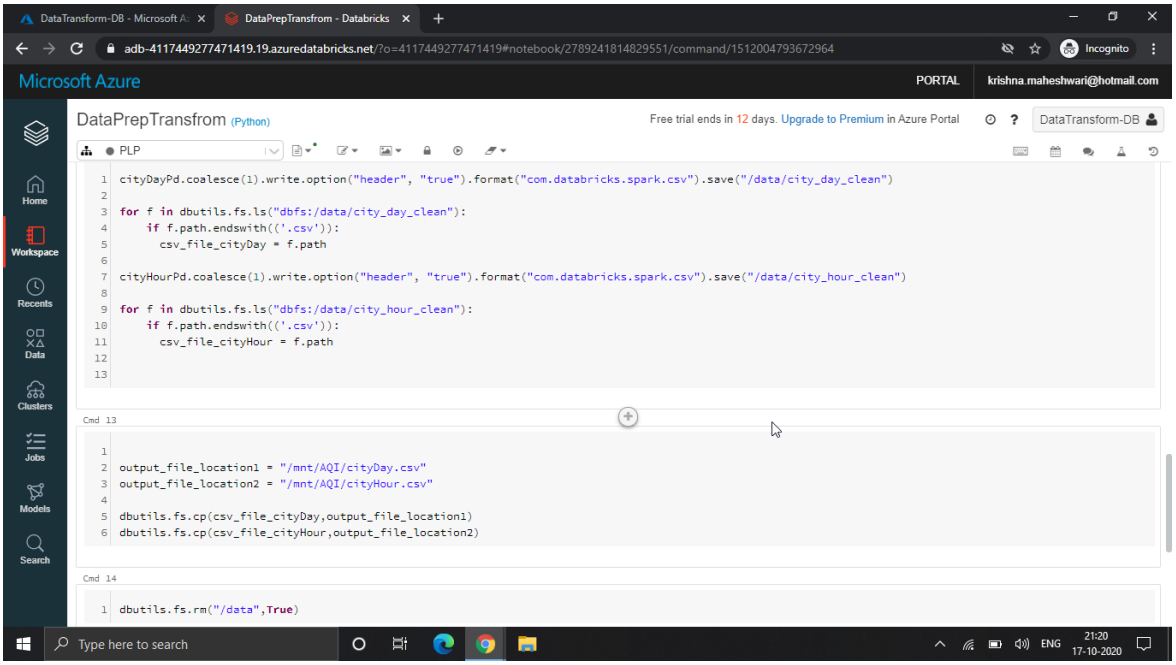
Creating Azure Databricks Instance



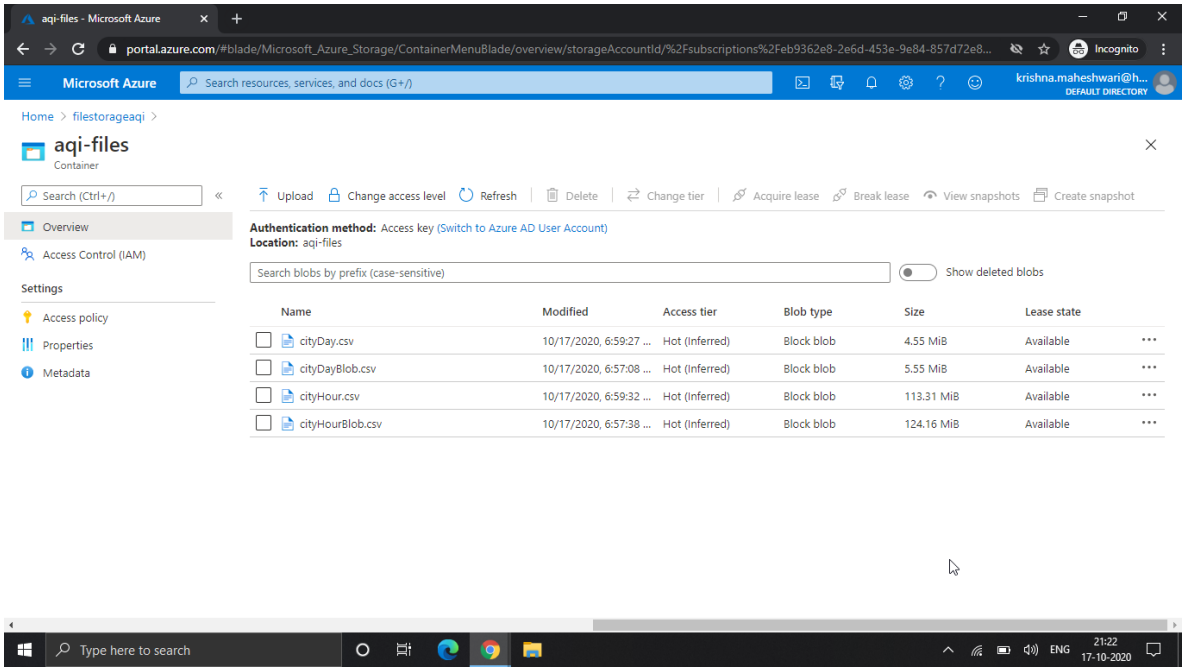
Generating Token



Saving Transformed Data To Blob



Verifying Data Transfer



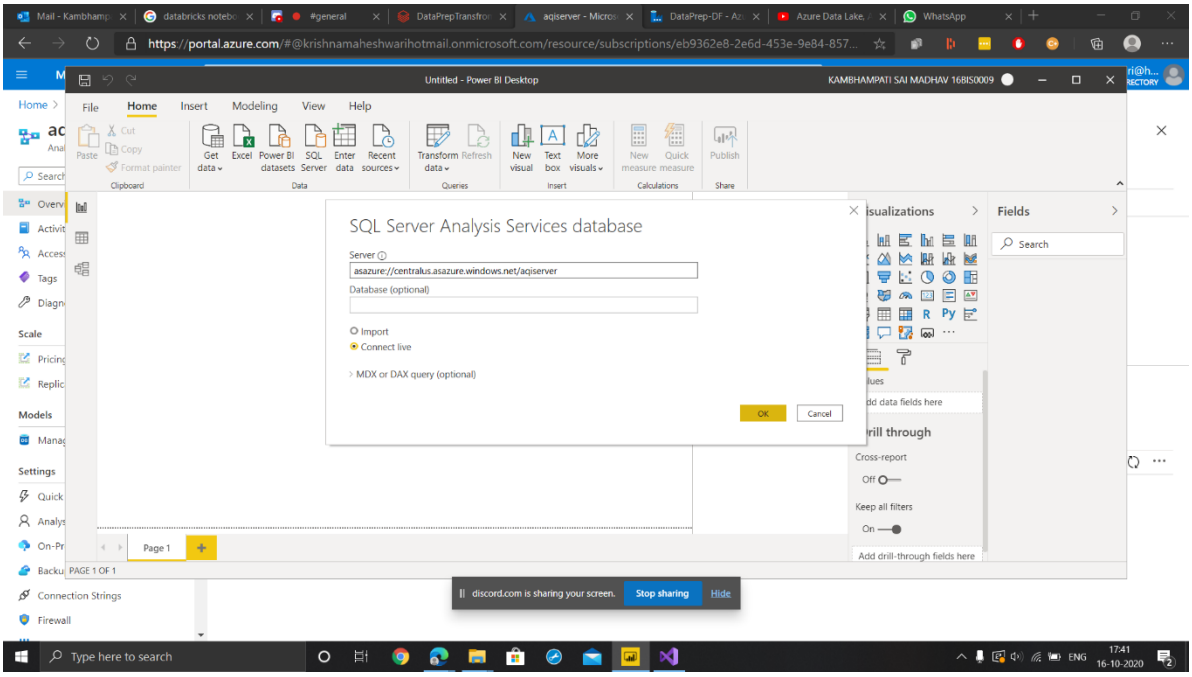
Data Model Deployment

Language Used – DAX

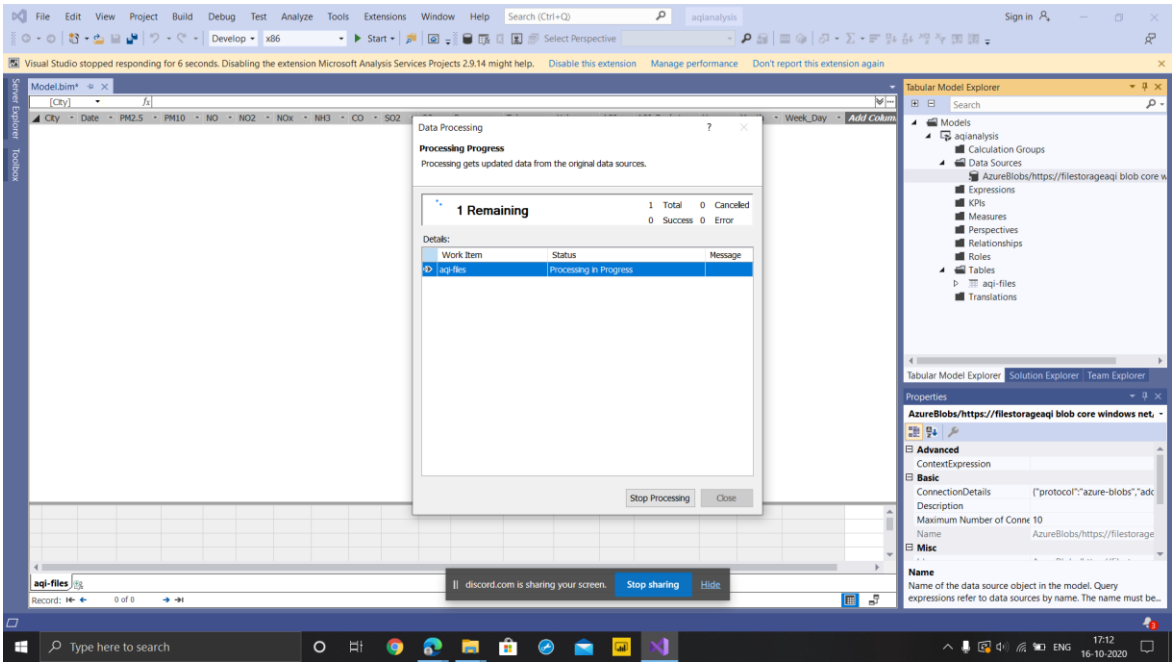
IDE Used – Visual Studio 2019

1. Importing Data Into Visual Studio Analysis Services Project
2. Establishing Internal Relationships Among Tables
3. Deploying Data Model To Azure Analysis Service Server

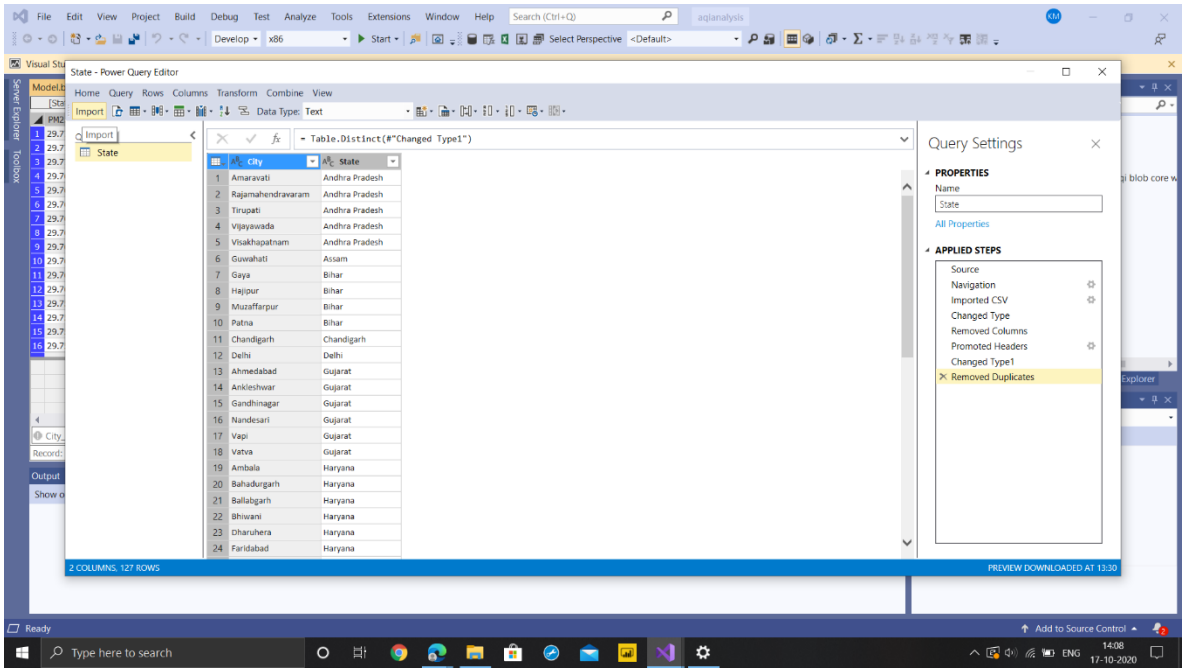
Connecting To Server



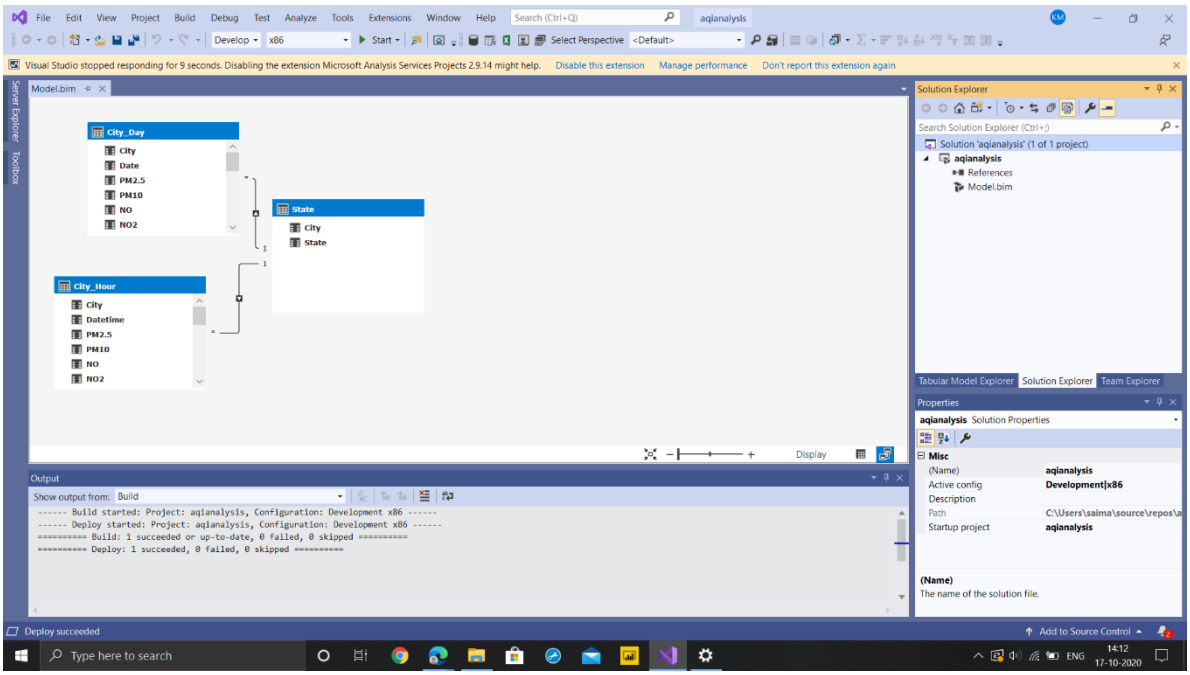
Connecting With Storage Account



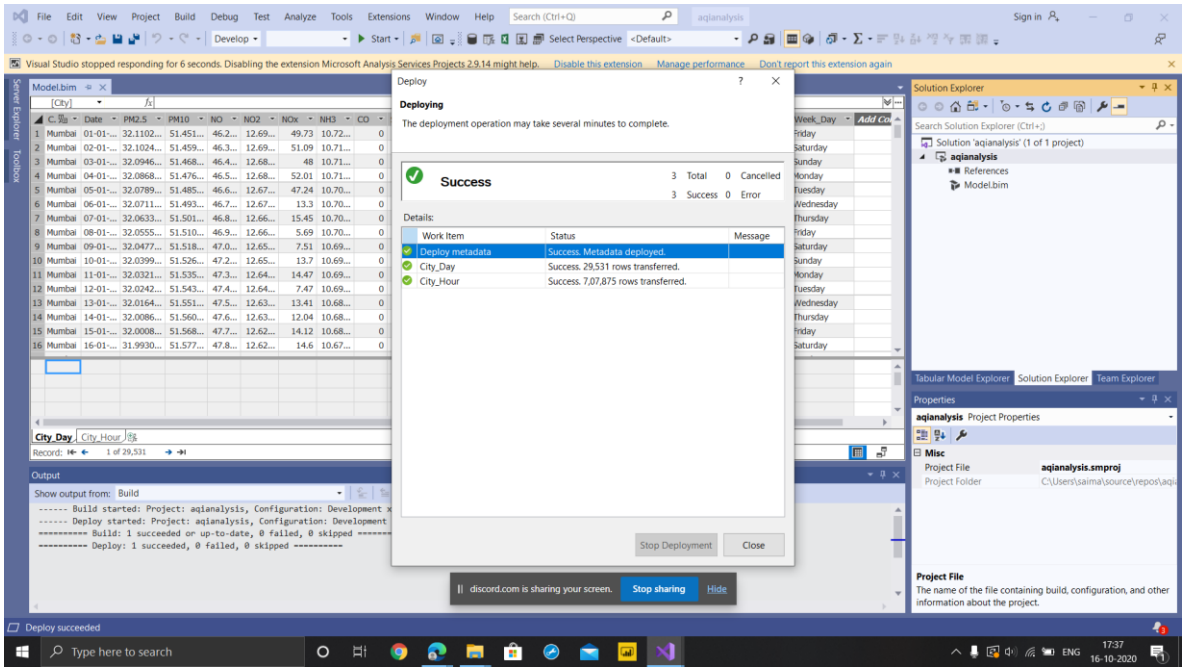
Fetching Data From Storage Account



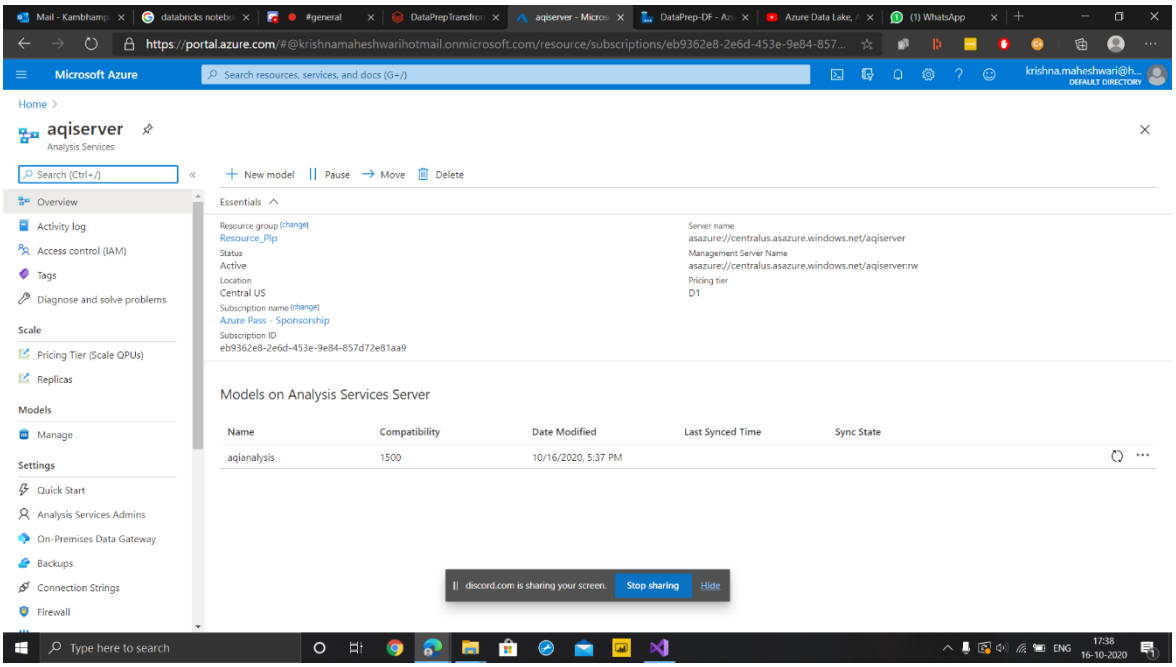
Establishing Internal Relationships Among Tables



Deploying Data Model In Server



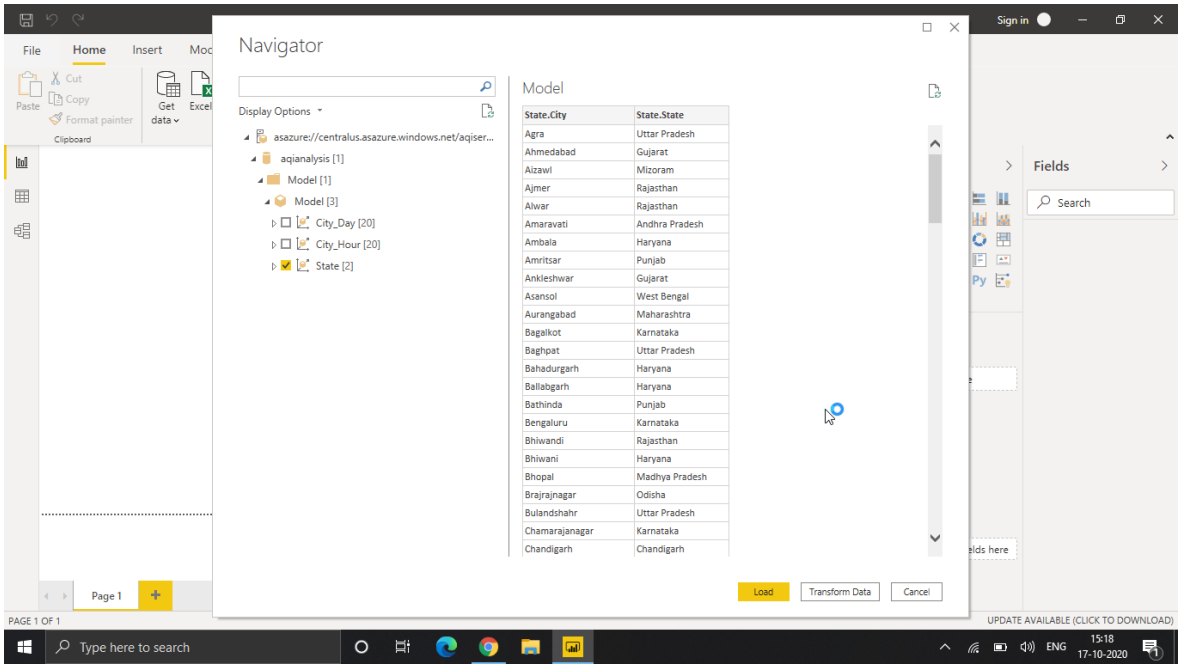
Verification In Analysis Server



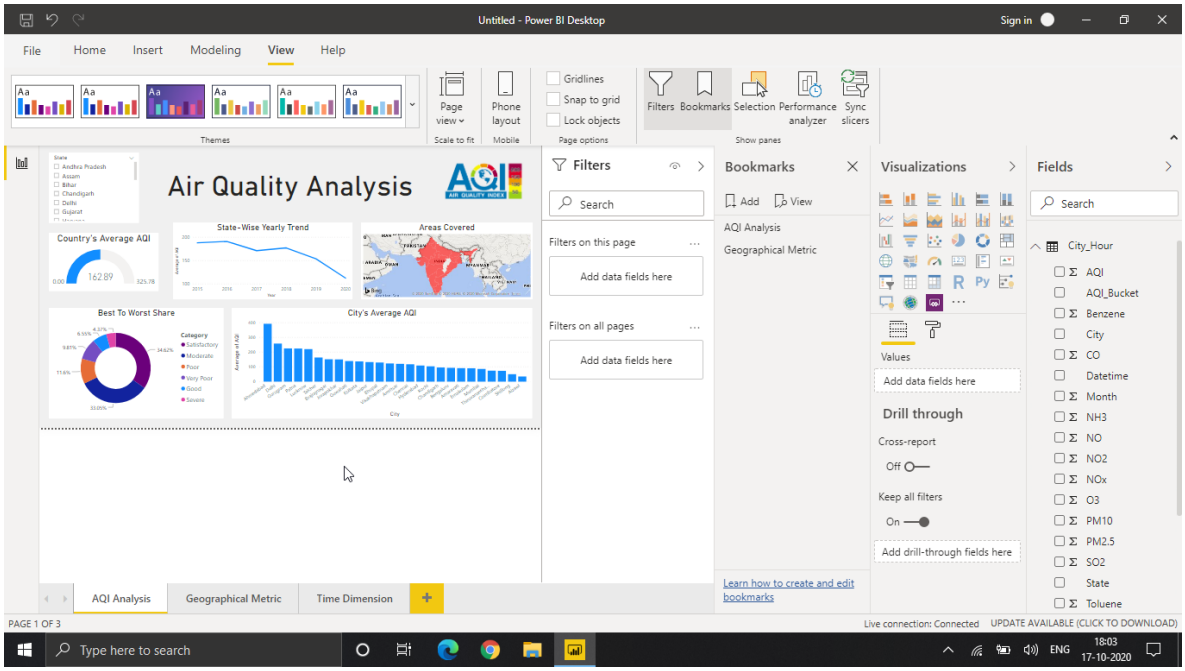
Data Analysis & Reporting

1. Connecting Power BI With Azure Analysis Service Server
2. Fetching Data Model From Server & Loading Data
3. Generating Reports Using Visualizations
4. Categorising Reports On Location & Time Basis
5. Covering Various Trends & Comparison Scenarios
6. Creating Various Slicers With Handled Interaction & Bookmarks

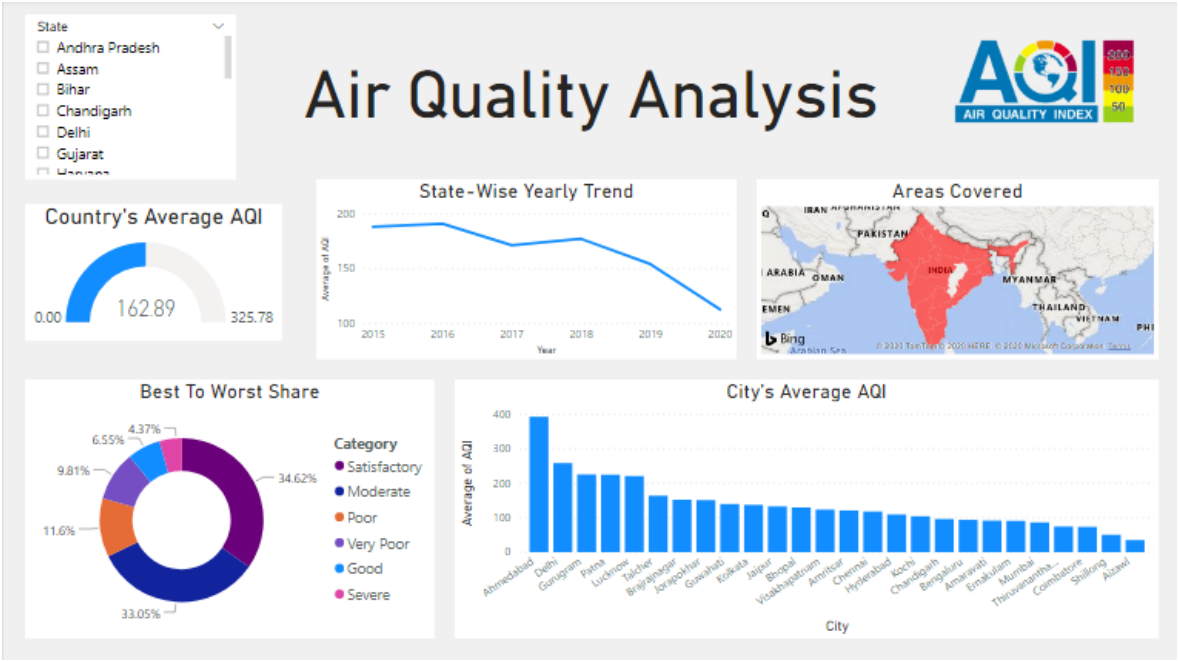
Connecting Power BI with Server



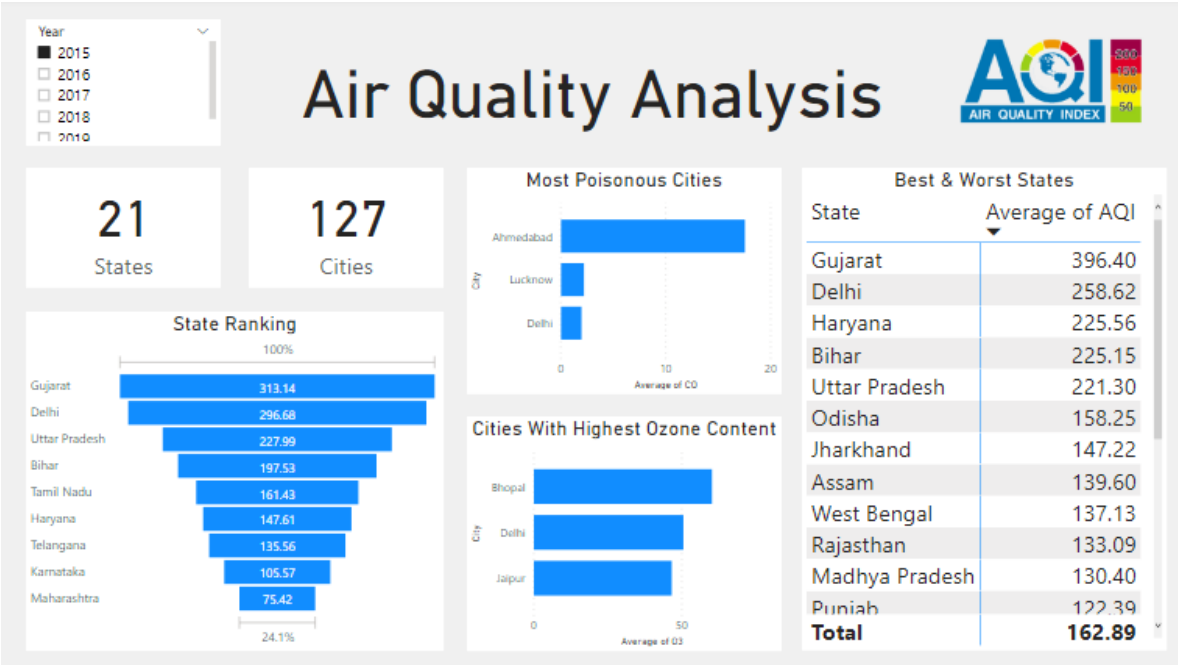
Creating Desired Slicers & Bookmarks



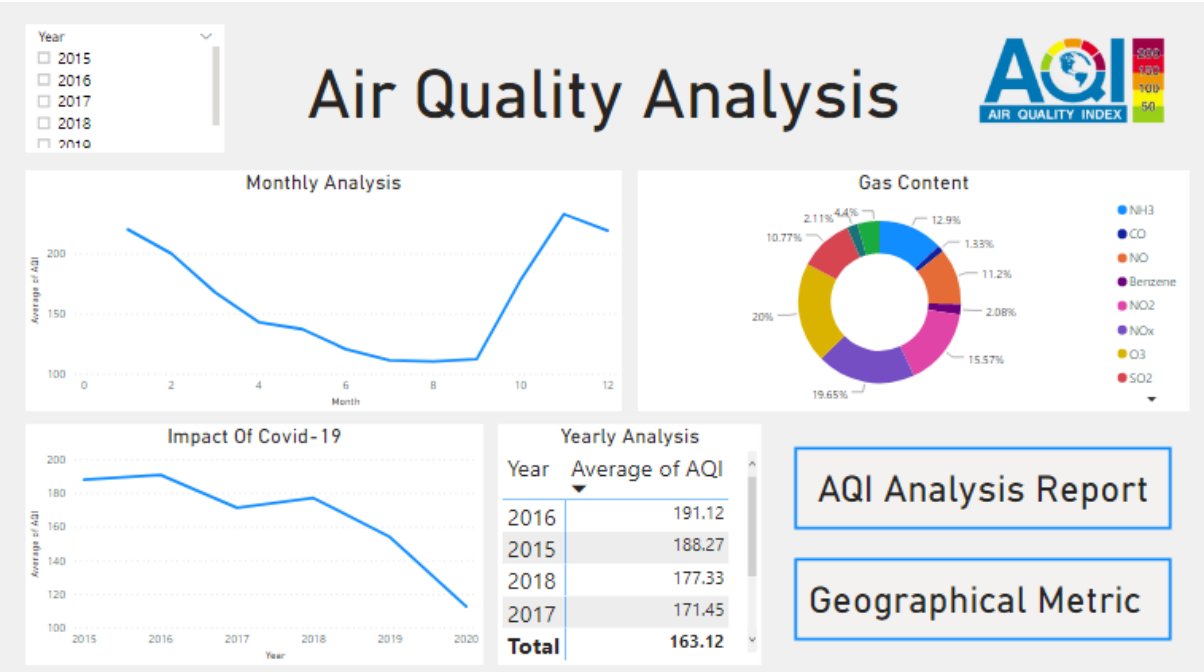
AQI Analysis



Geographical Metric



Time Dimension



- - End - -