**DATA SCIENCE PROJECT REPORT**

(Project Semester August-December 2019)

***(AIR QUALITY CHECKS)***

Submitted by

Ankit yadav

Registration No. 11707994

Integrated B.Tech CSE – MBA

Course Code: KM094

Under the Guidance of

**Ms. Vasudha**

**Assistant Professor**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Ankit Yadav bearing Registration no. 11707994 has completed KM094 – Introduction to Data Management project titled, **“Air quality checks.”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, Ankit yadav student of Integrated B.Tech CSE - MBA under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:

Signature

Registration No. 11707994 Name

**Contents**

|  |  |  |
| --- | --- | --- |
| Serial No. | Topic | Page No. |
| 1 | Introduction | 5 |
| 2 | Scope of Analysis of Project using Tableau Prep | 6 |
| 3 | Analysis of Data through Excel | 13 |
| 4 | Source of the data set | 18 |
| 5 | Analysis Done | 19 |
| 6 | Future Scope | 20 |
| 7 | Bibliography | 20 |

**INTRODUCTION**

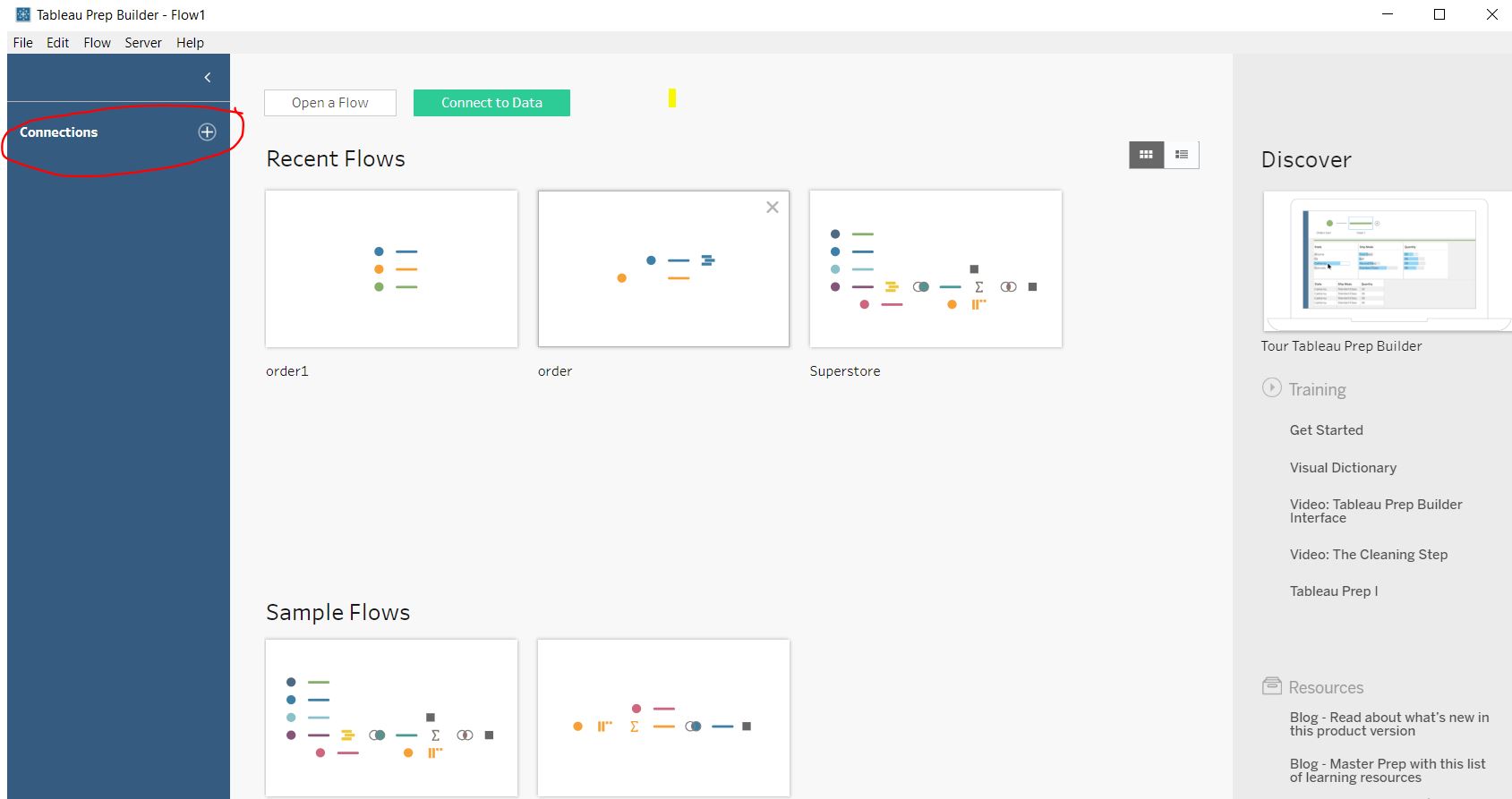
With rapid innovations and surge of Internet companies like google, Yahoo, Amazon, Ebay and a rapidly growing internet savvy population, today’s advanced systems and enterprises are generating data in a huge volume with great velocity and in a multi -structured formats including videos, images, sensor data, weblogs, etc. from different sources.

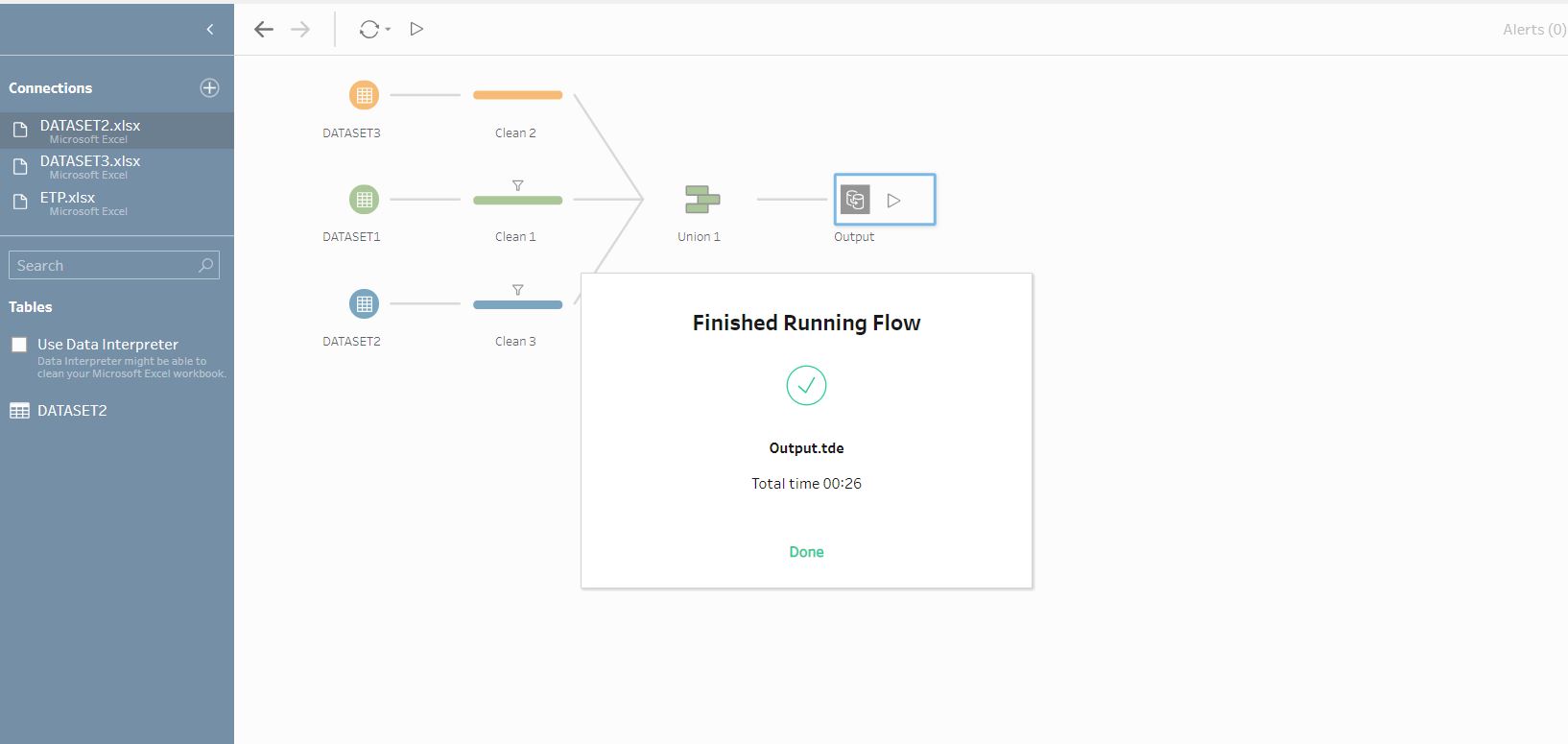
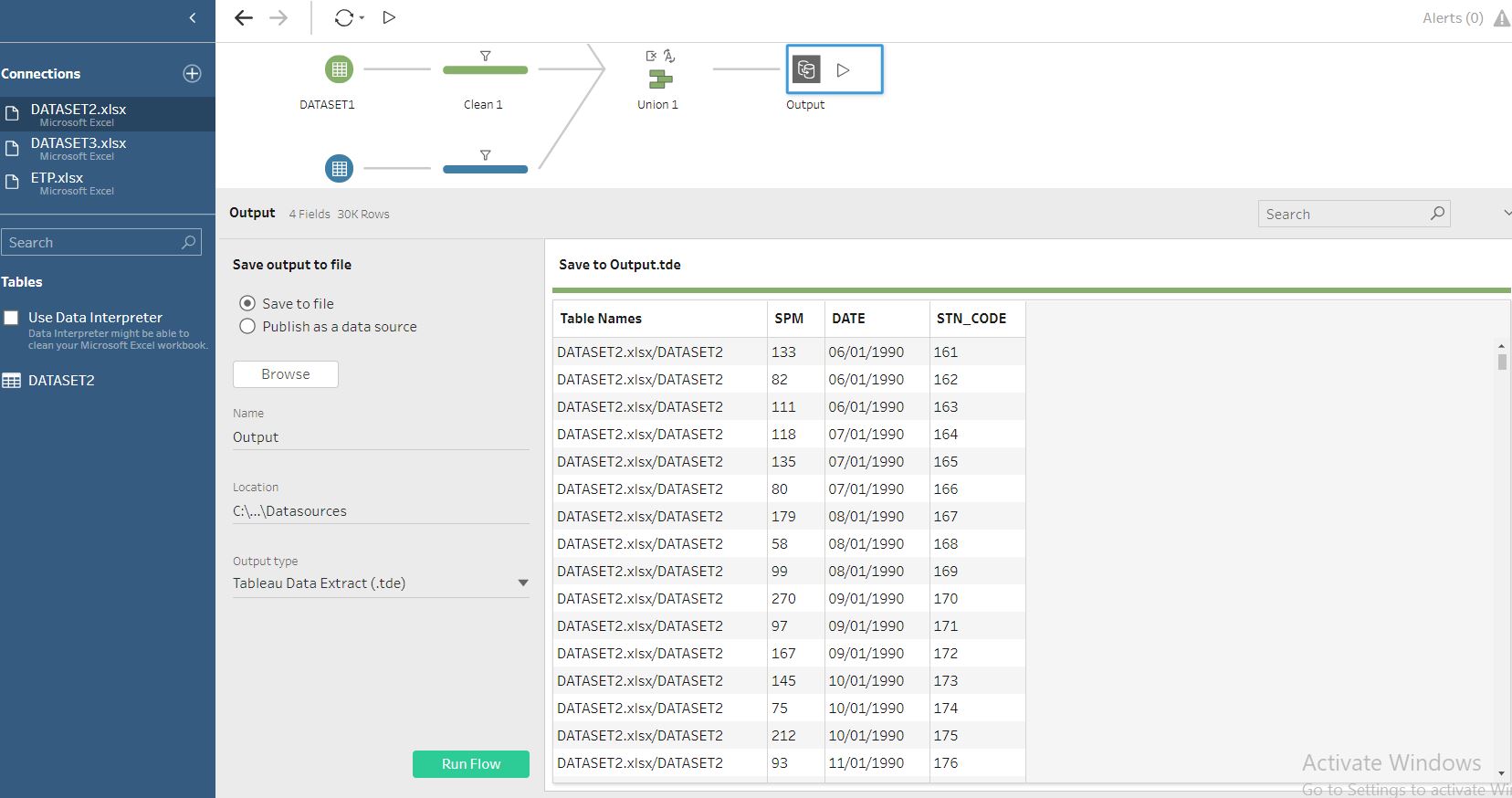
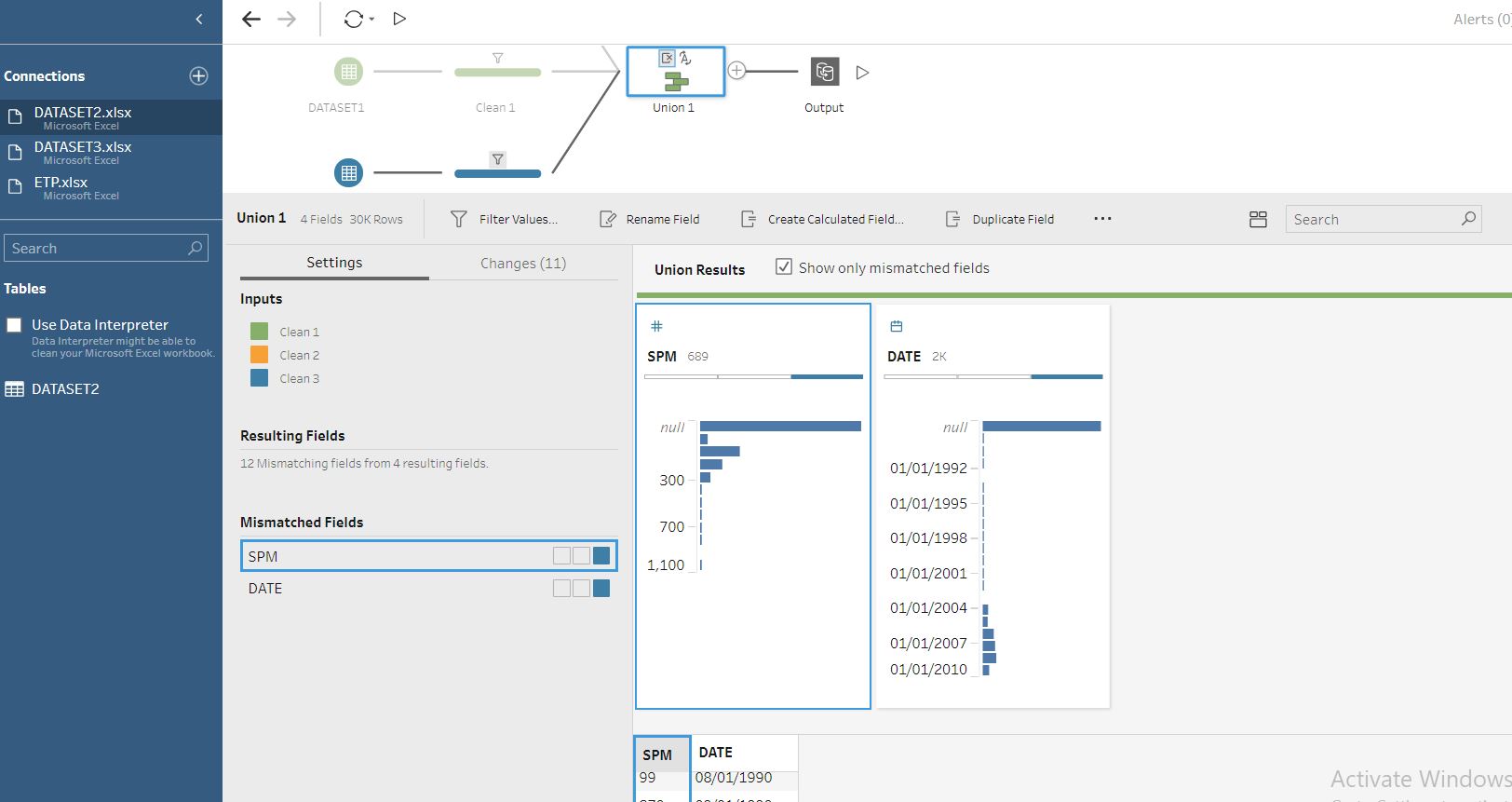
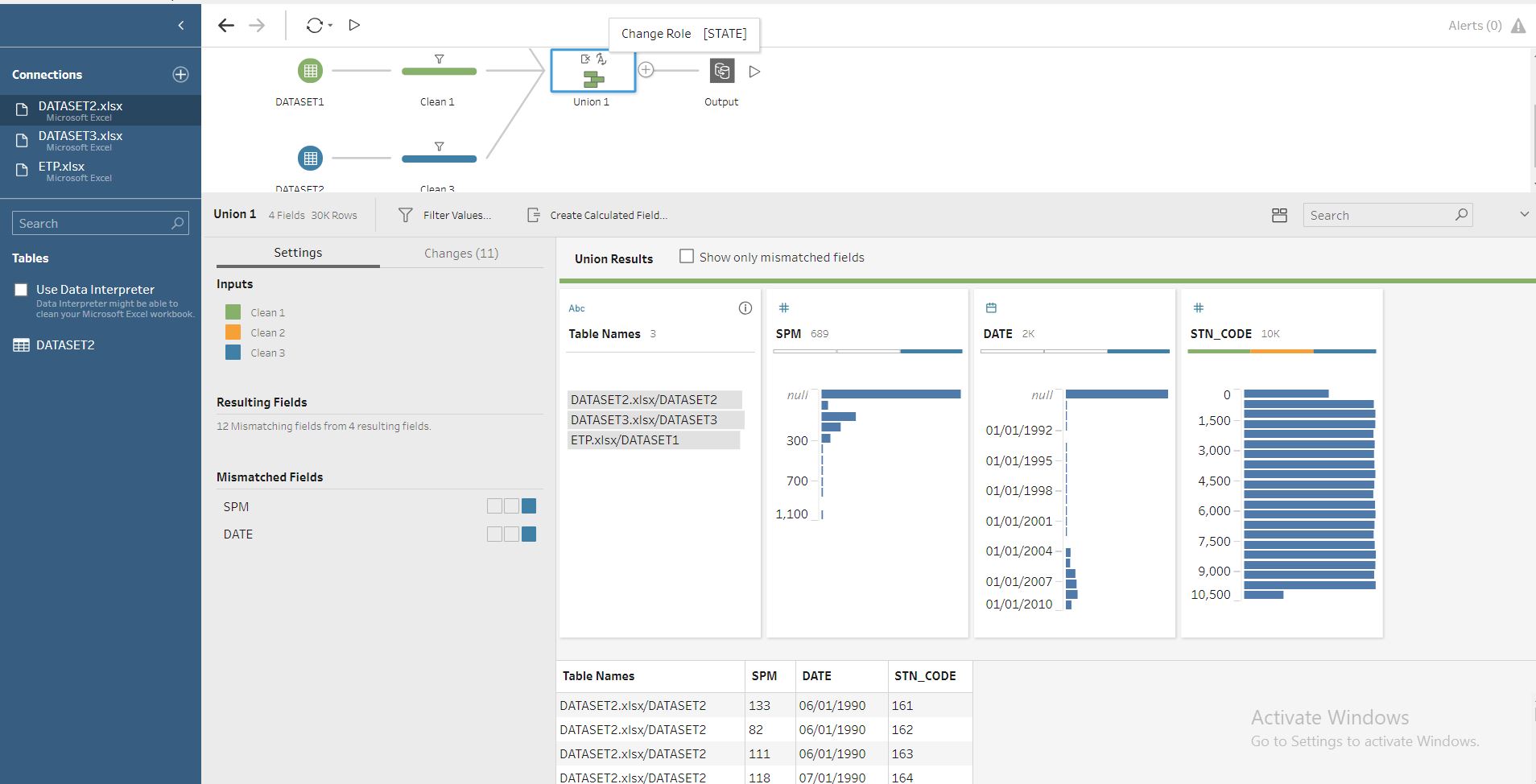
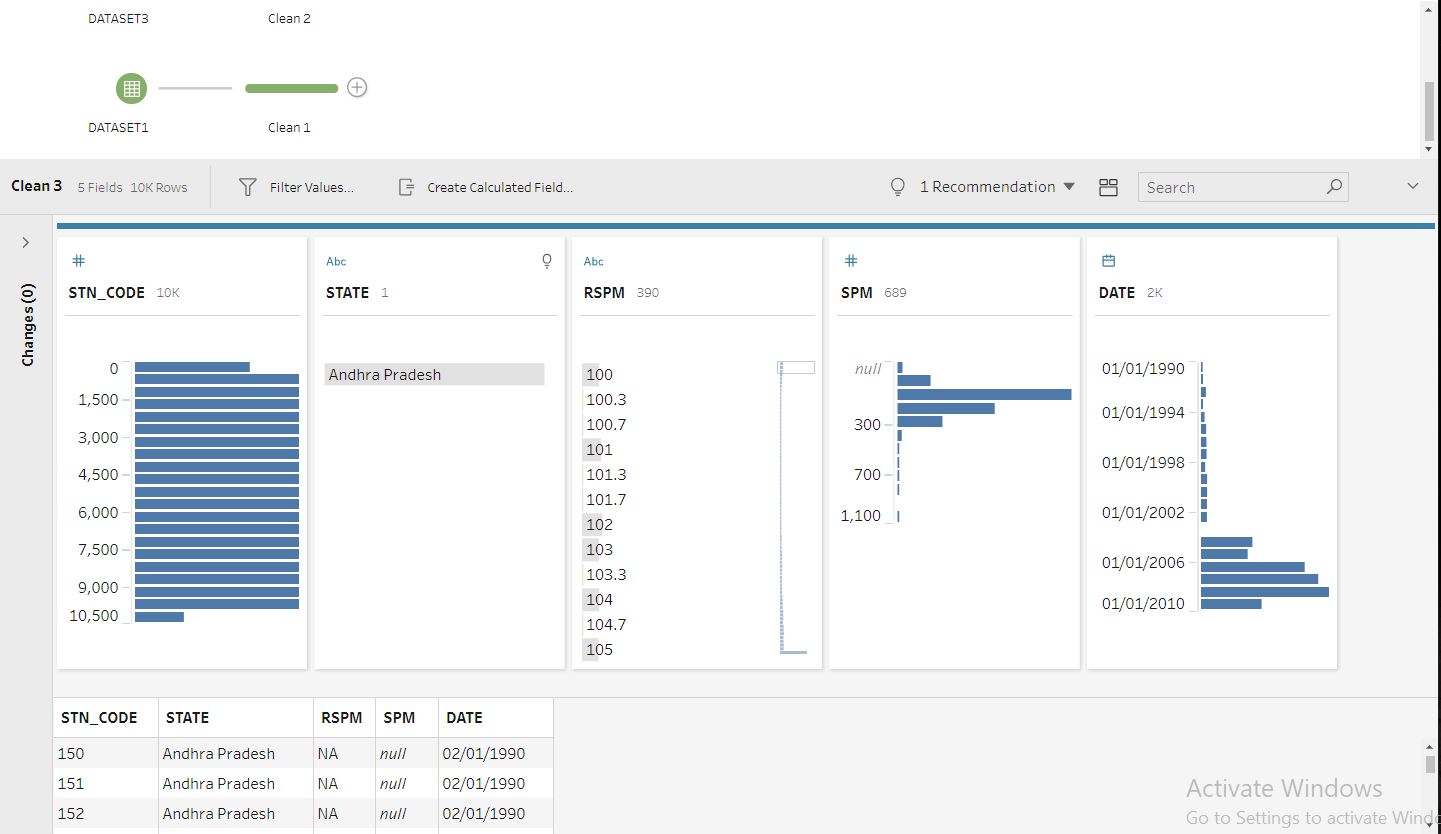
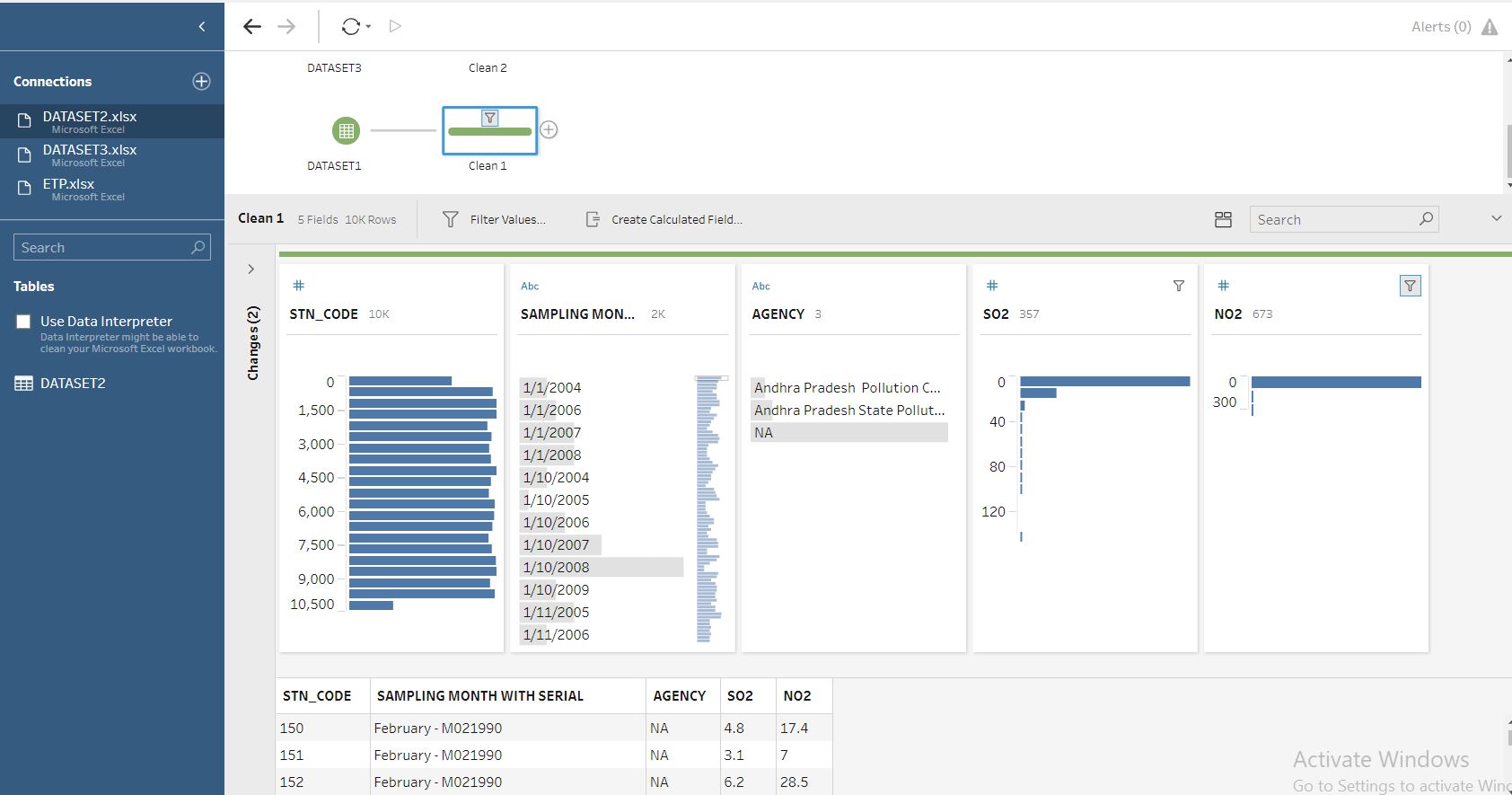
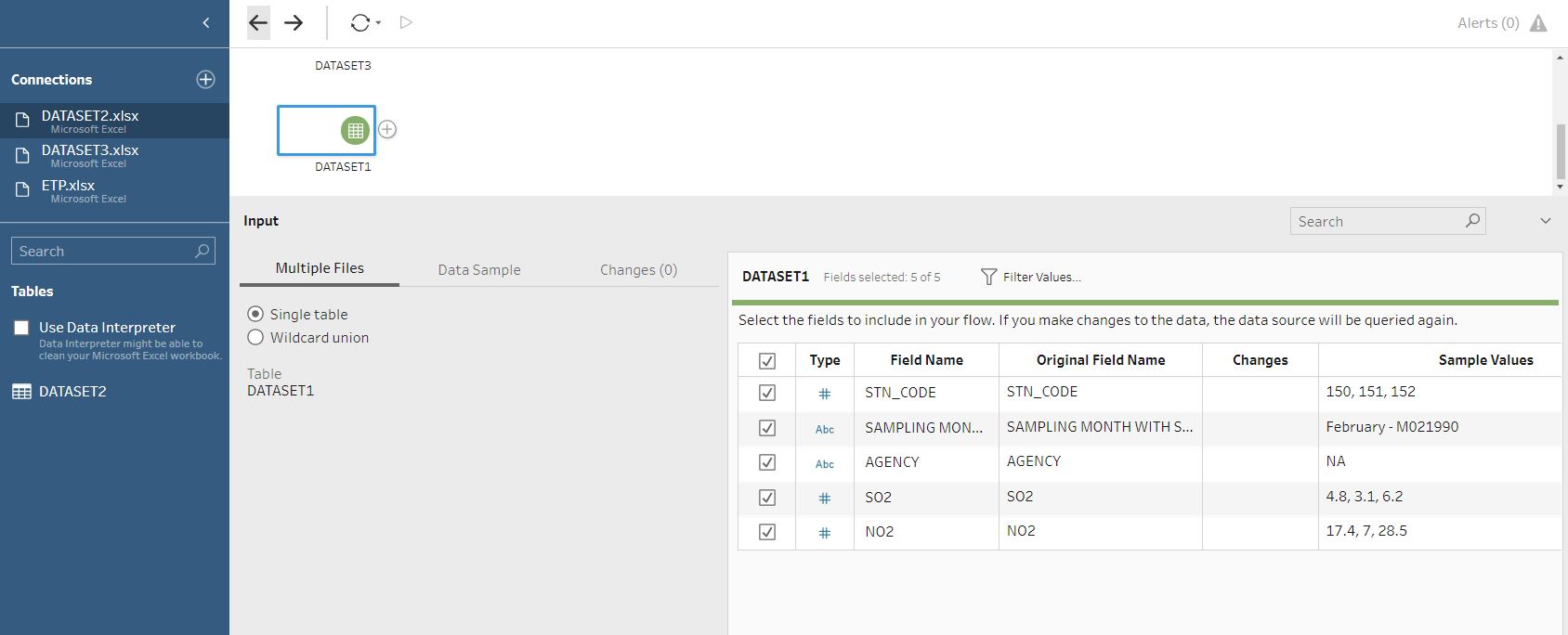
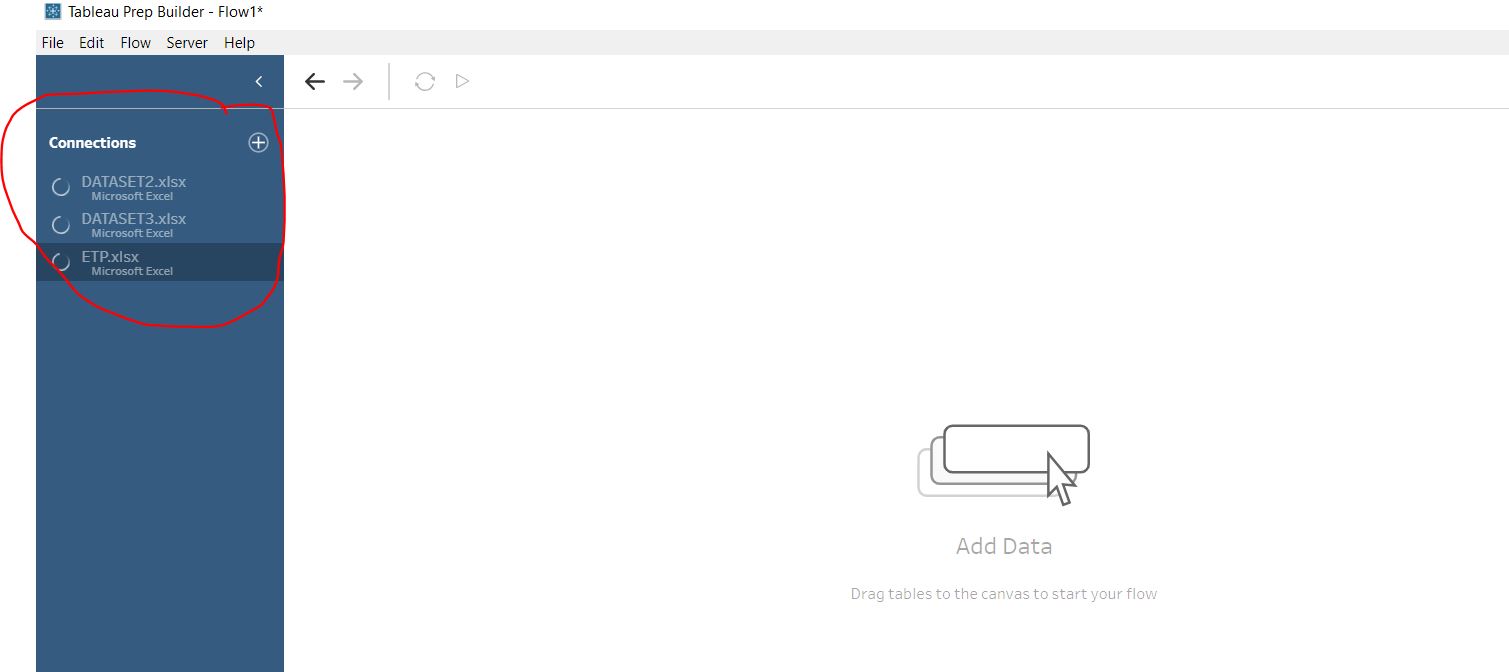
Data science is a multifaceted discipline, which encompasses machine learning and other analytic processes, statistics and related branches of mathematics, increasingly borrows from high performance scientific computing, all in order to ultimately extract insight from data and use this new-found information to tell stories. Data science is both synonymous with data mining, as well as a superset of concepts which **includes** data mining.

In this Specialization learners will develop foundational Data Science skills to prepare them for a career or further learning that involves more advanced topics in Data Science. The specialization entails understanding what is Data Science and the various kinds of activities that a Data Scientist performs. It will familiarize learners with various open source tools, like Jupyter notebooks, used by Data Scientists. It will teach you about methodology involved in tackling data science problems. The specialization also provides knowledge of relational database concepts and the use of SQL to query databases. Learners will complete hands-on labs and projects to apply their newly acquired skills and knowledge.

Every Specialization includes a hands-on project. You'll need to successfully finish the project(s) to complete the Specialization and earn your certificate. If the Specialization includes a separate course for the hands-on project, you'll need to finish each of the other courses before you can start it.

Scope of Analysis of Project using Tableau Prep.

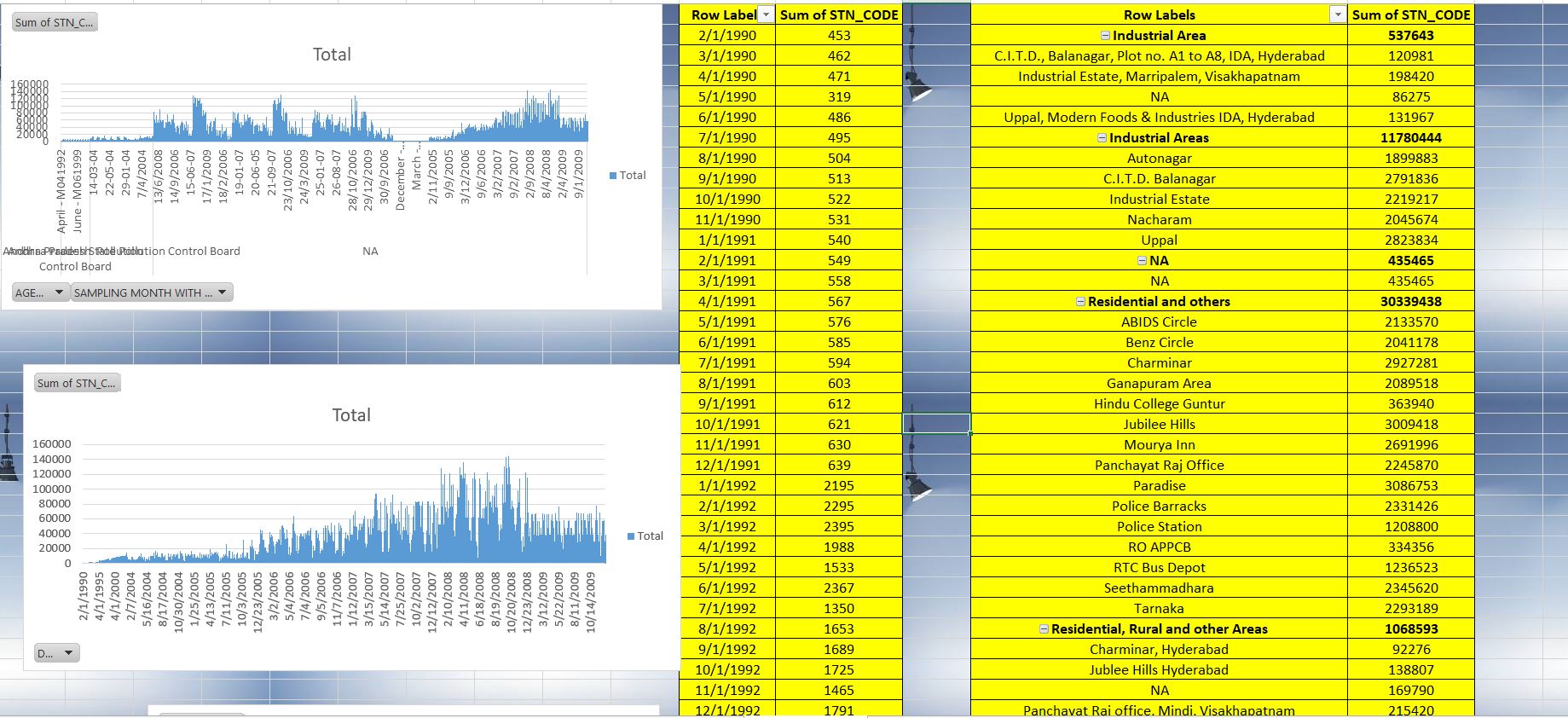
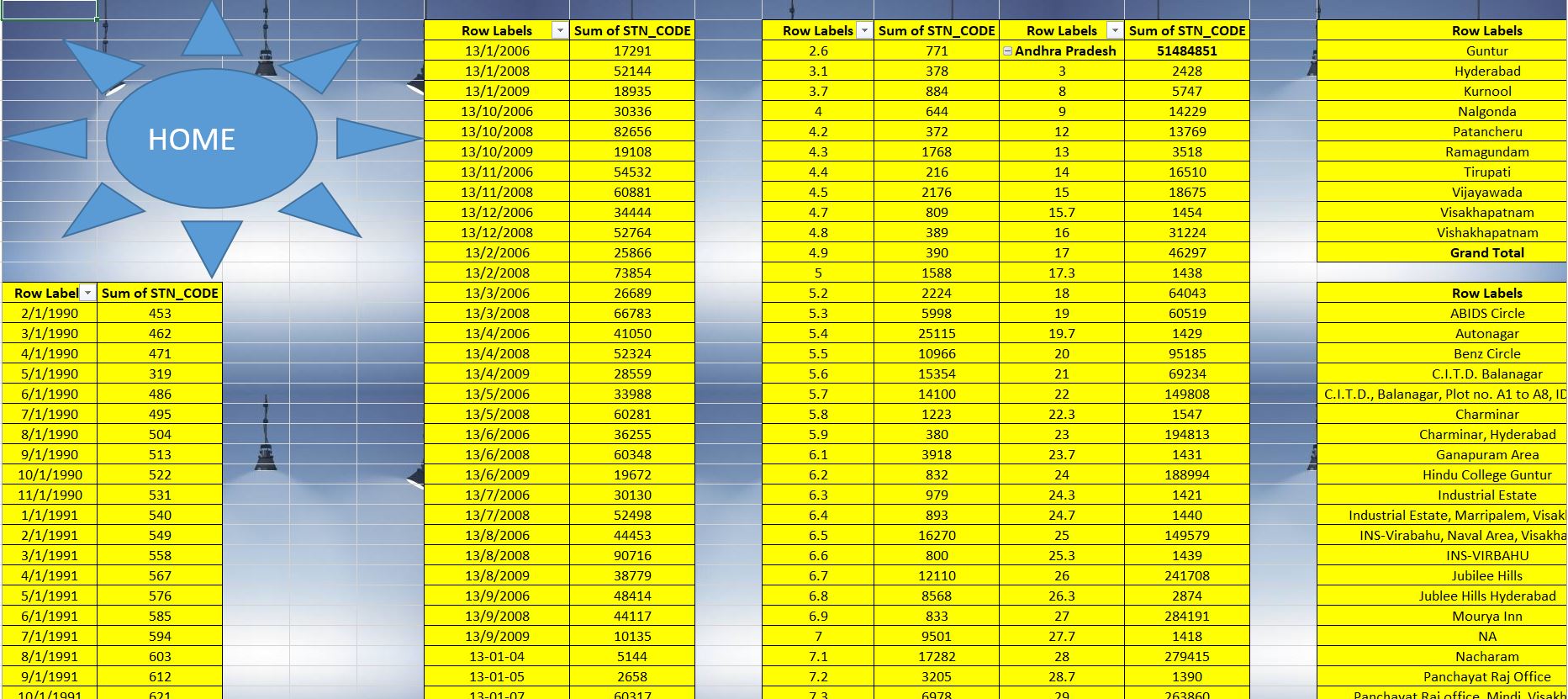
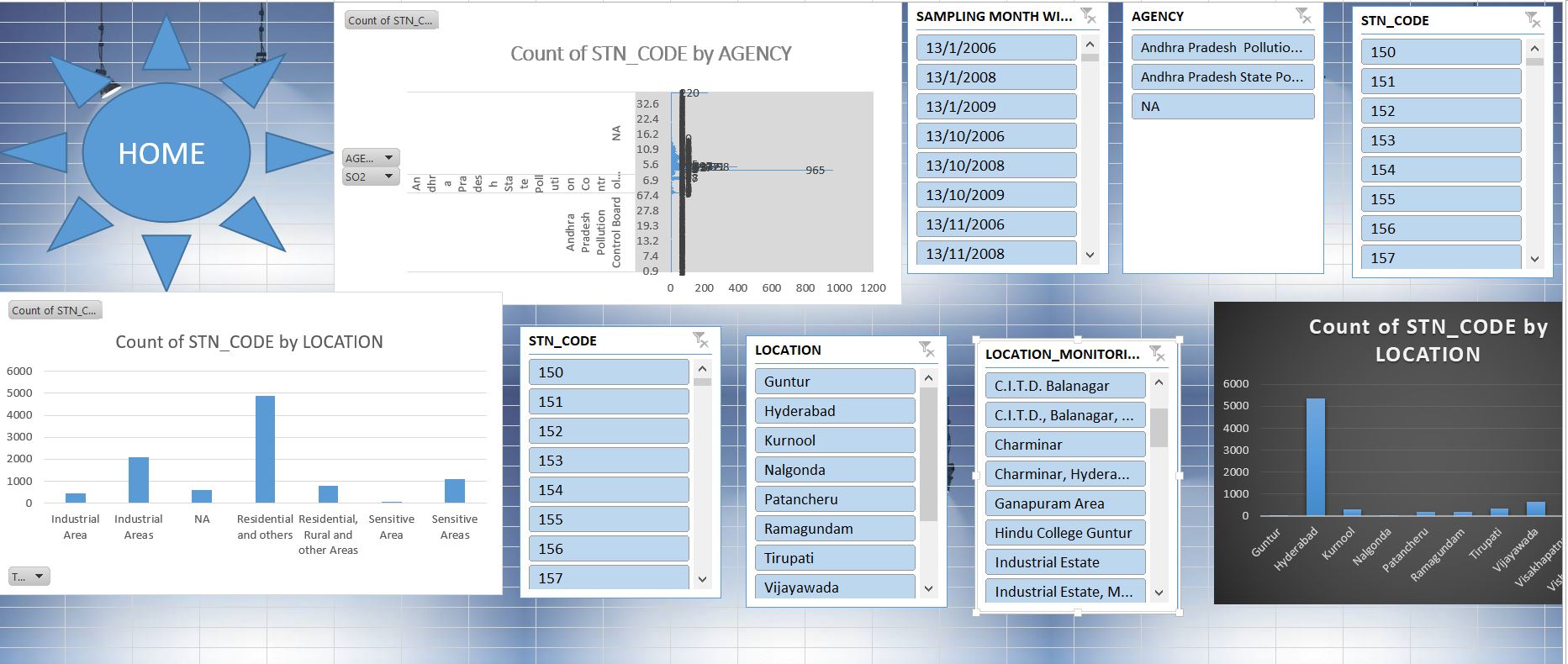
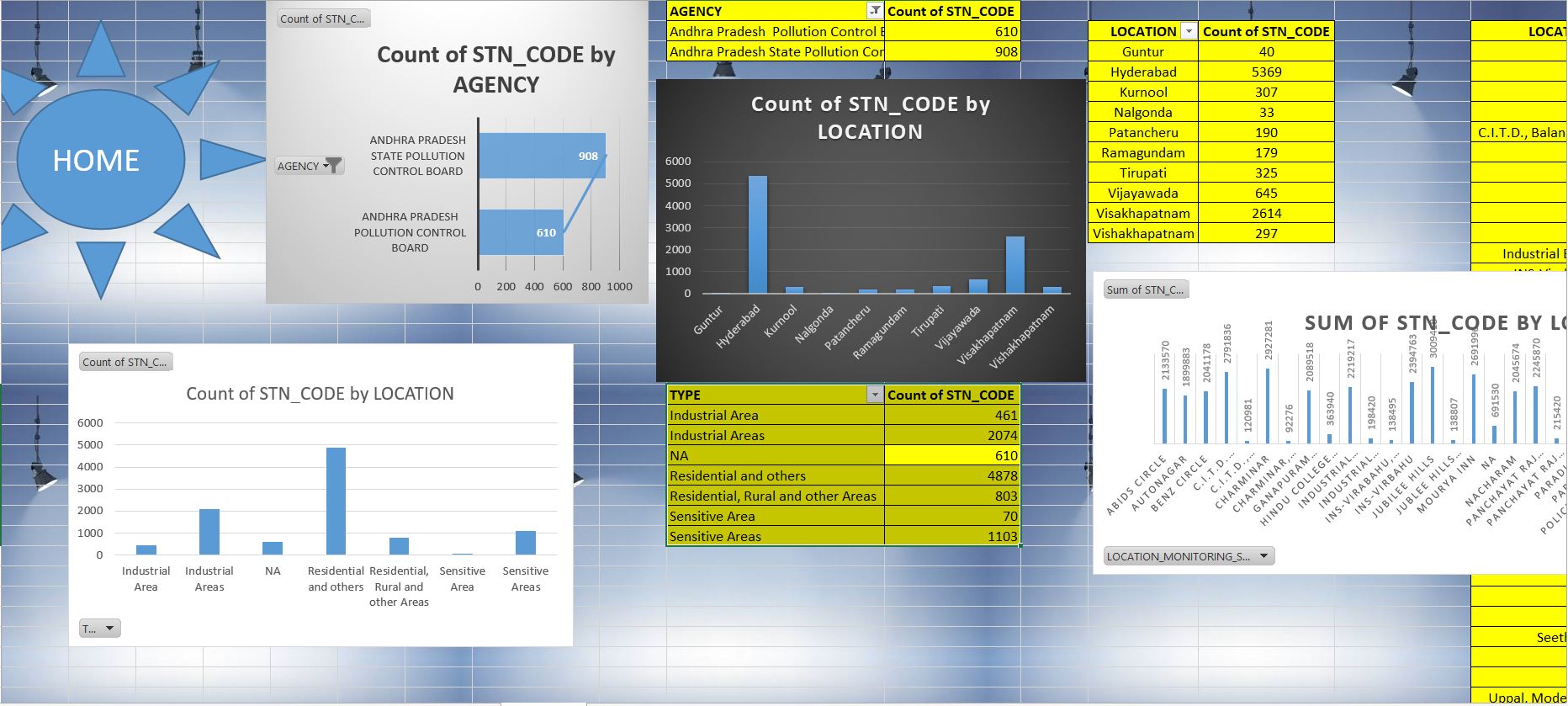
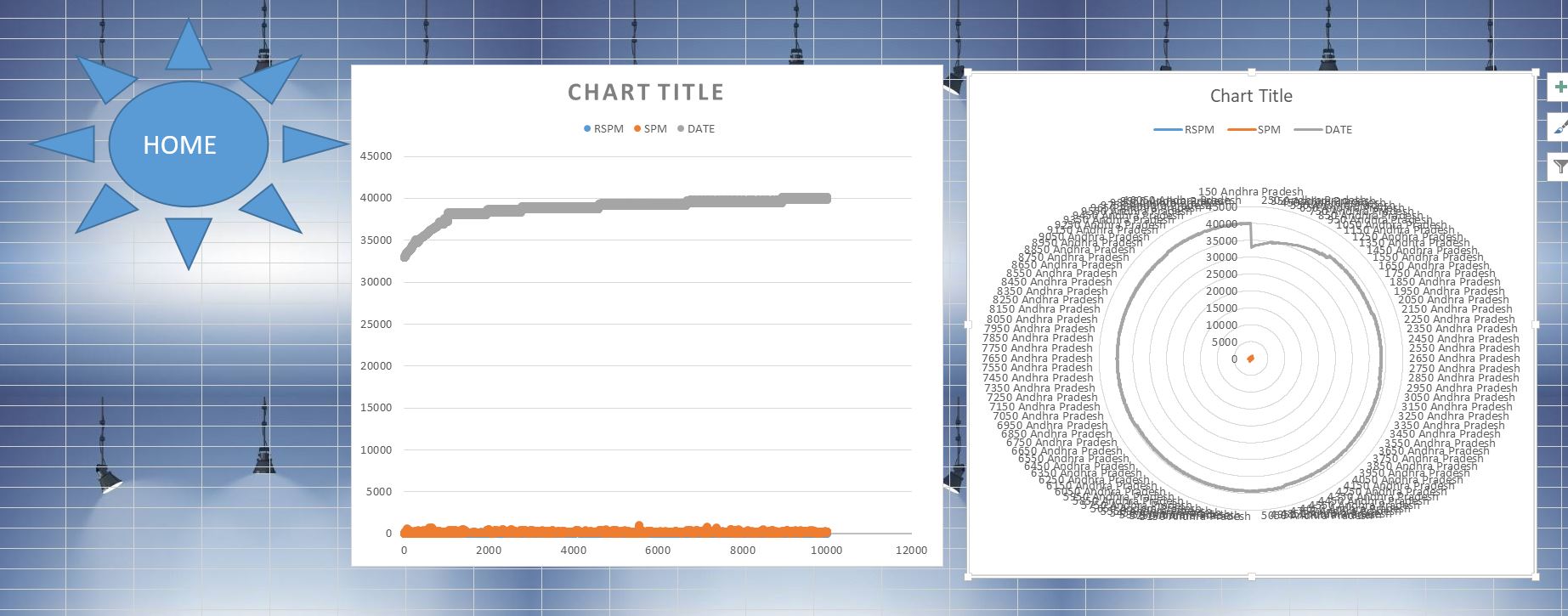
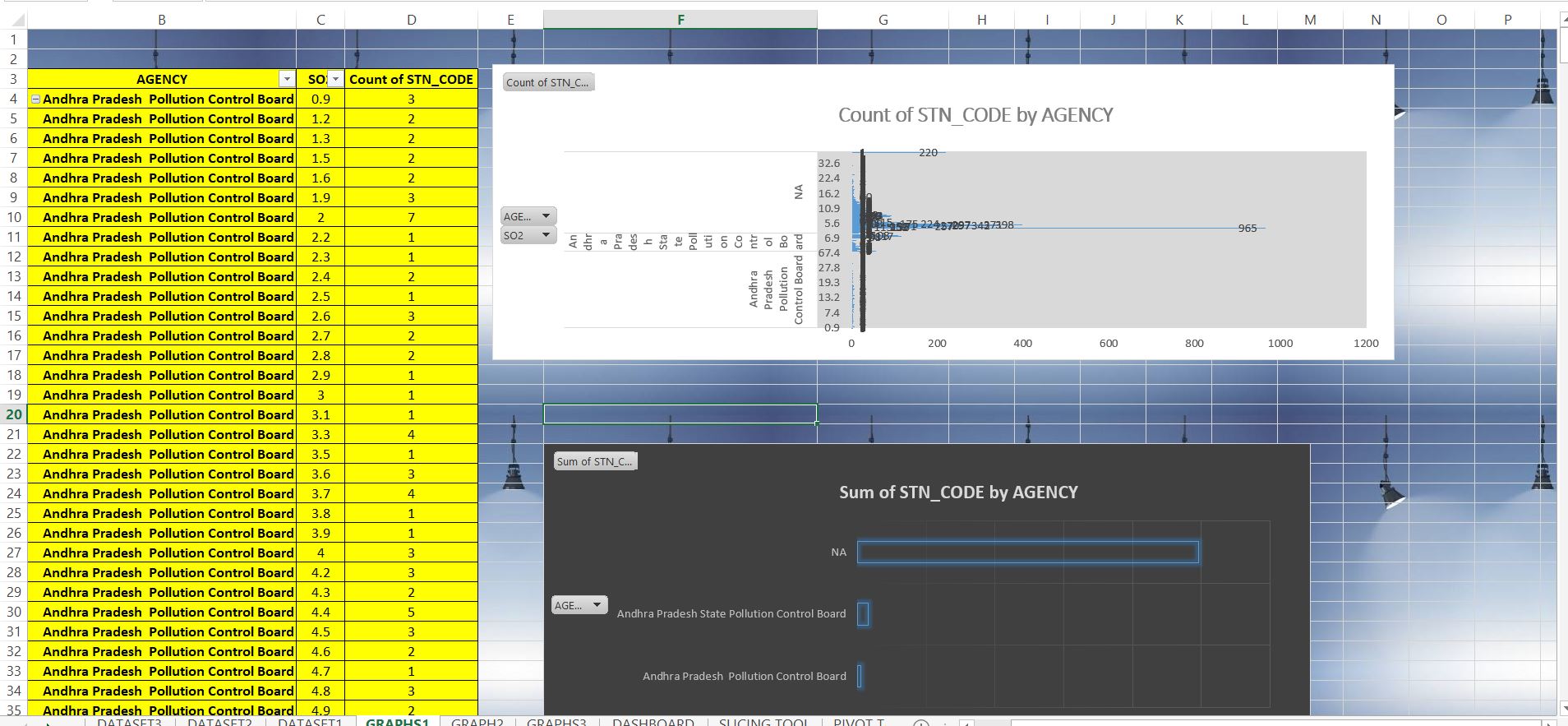
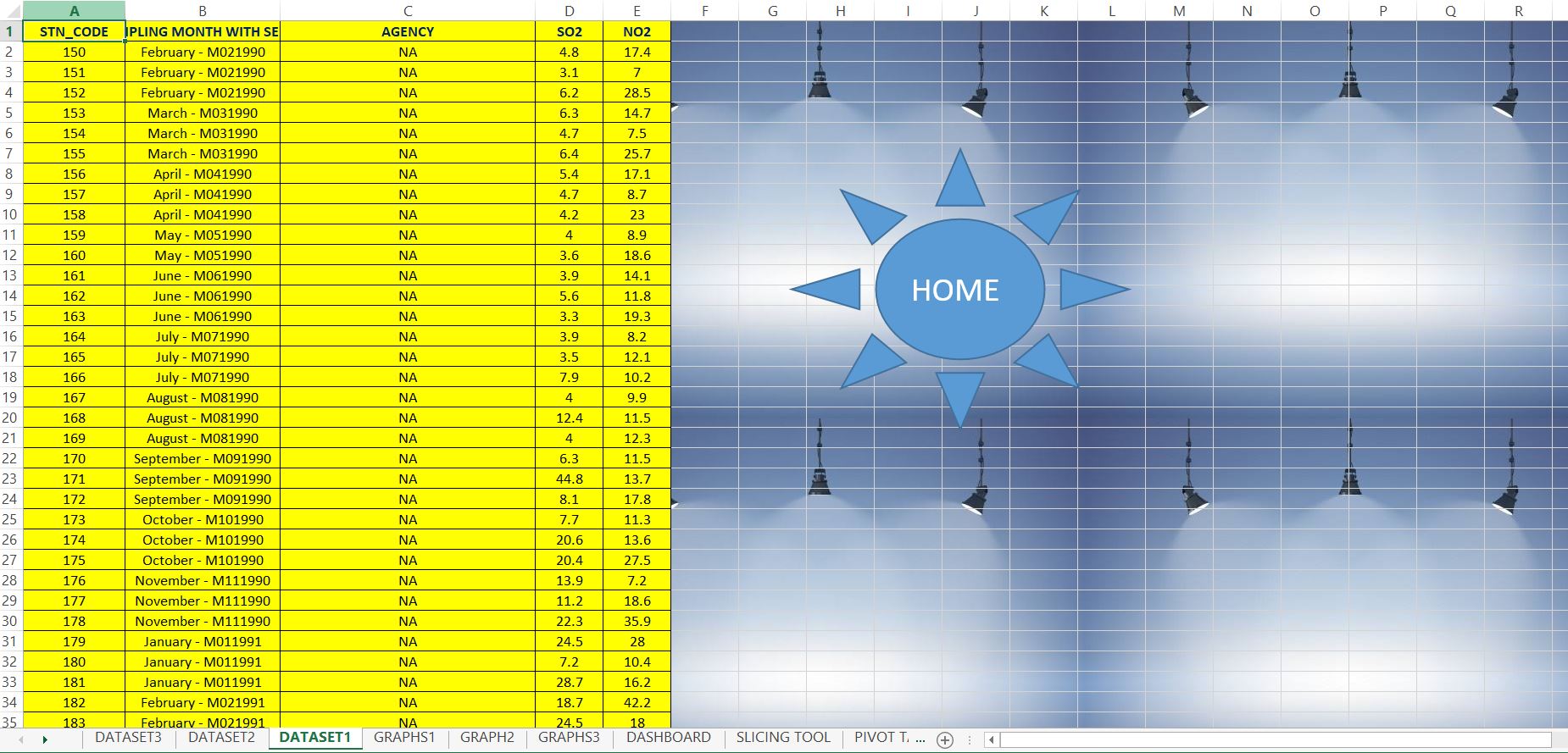
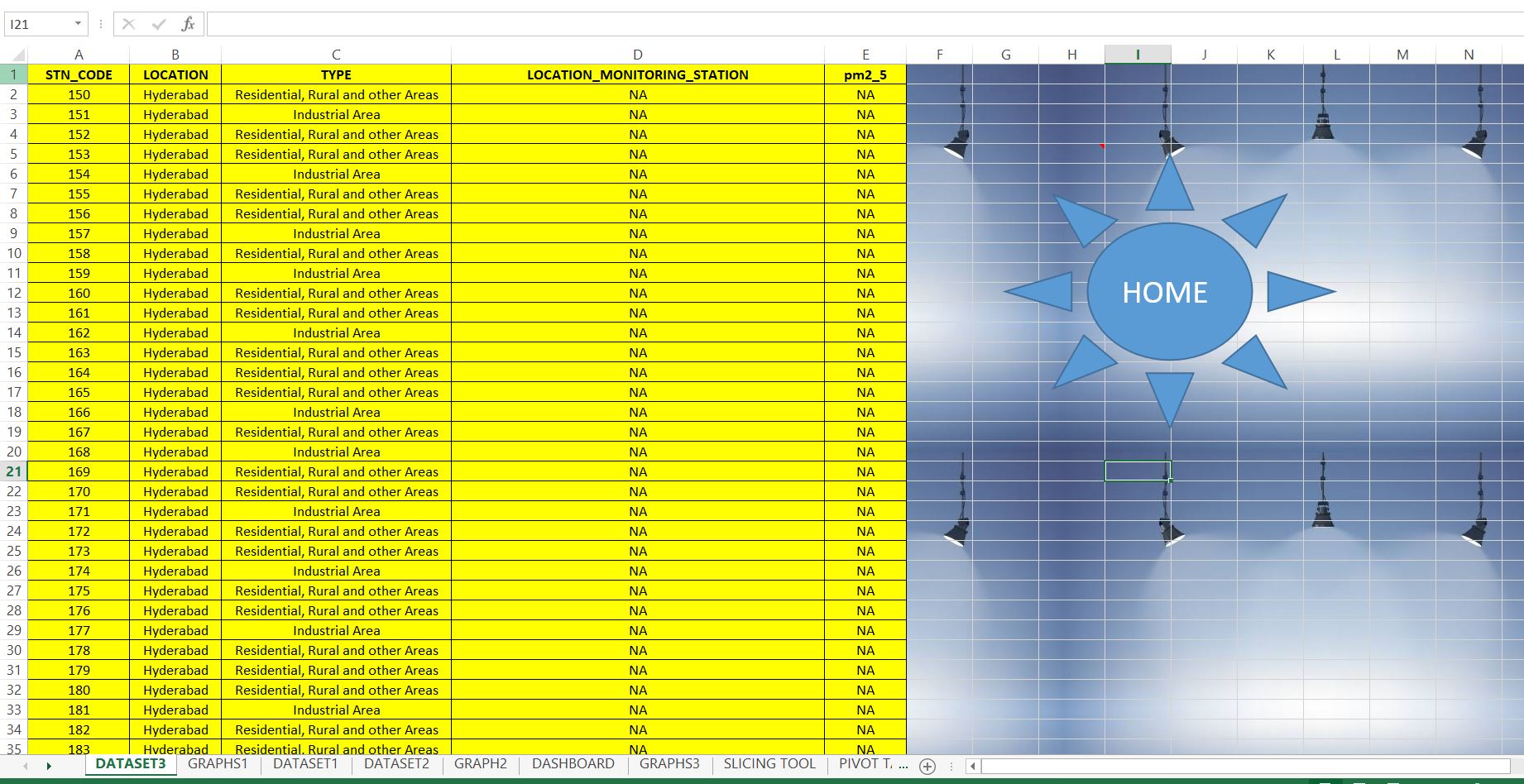
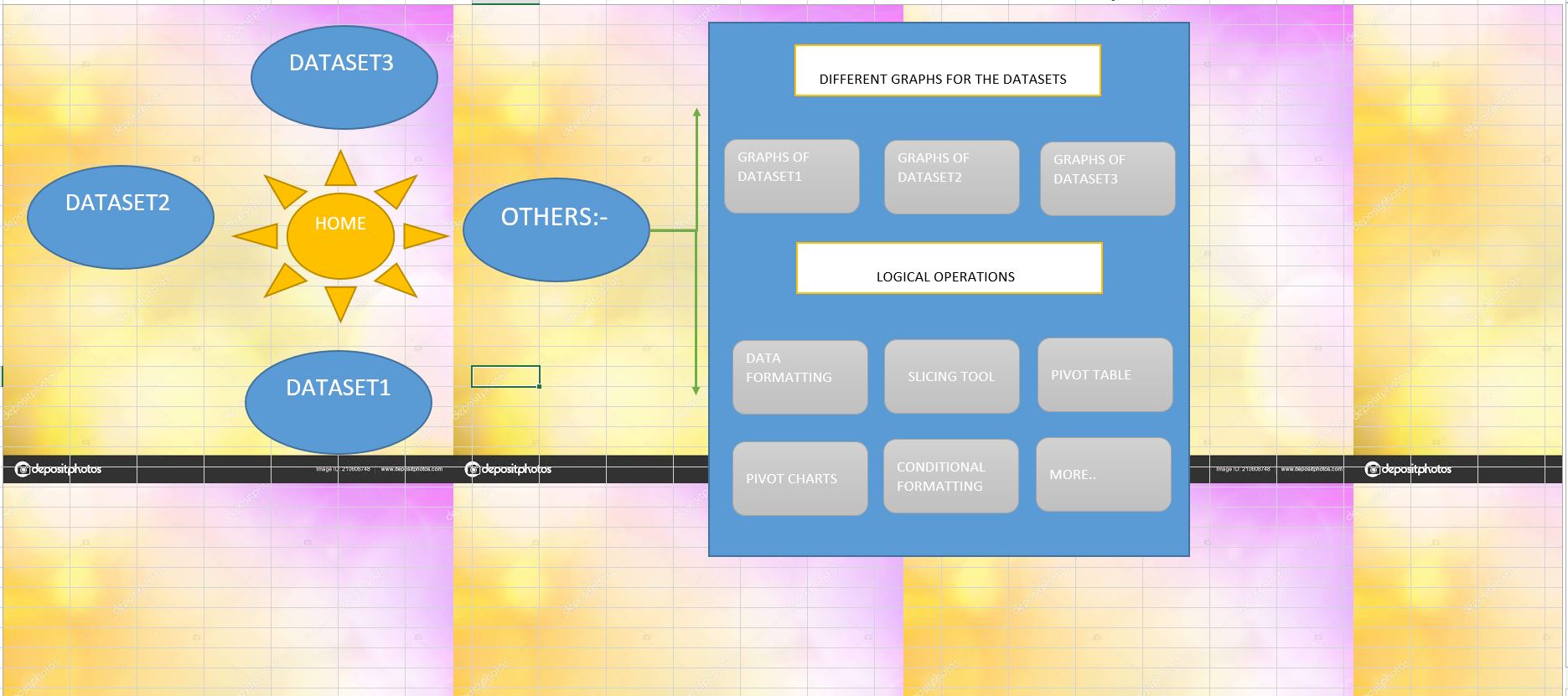




|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated |  |

Analysis of Data through Excel:

Dashboard:



**Source of the data set:**

Some of the major data set for analysis can be obtained from open source data base like Data.gov.in , Datacamp ,kaggle.

The data which is used in this analysis is downloaded from kaggle platform.

Some of link of website where you can get the dataset are given below:-

<https://catalog.data.gov/dataset>

<https://toolbox.google.com/datasetsearch>

<https://www.kdnuggets.com/datasets/index.html>

<https://archive.ics.uci.edu/ml/datasets.php>

<https://www.datacamp.com/?utm_source=adwords_ppc&utm_campaignid=1242944157&utm_adgroupid=58673827368&utm_device=c&utm_keyword=datacamp&utm_matchtype=e&utm_network=g&utm_adpostion=1t1&utm_creative=340731356764&utm_targetid=kwd-297372810188&utm_loc_interest_ms=&utm_loc_physical_ms=20466&gclid=CjwKCAiA_MPuBRB5EiwAHTTvMa1y8NSVn0oGrH_id5PqNxiyDArf_0S6snV5bItuimKCV_VRh1JMjxoCmIwQAvD_BwE>

<https://www.kaggle.com/>

**Analysis Done:**

Various Method used in this analysis are as follows:-

Data Cleaning using Tableau Prep

Data validation

Slicing

Pivot Table

Pivot Chart

Visualization

Dashboard

Hyperlinking

Logical Operations

Cell formatting

Text formatting

Functional operation

Data handling etc.

**Future Scope:**

Future Scope of this project is that it predict the increase in level of toxicity in air at different location like states and cities .By this we can have a ideas that this much amount of increase will take place so we can take immediate actions to prevent it and to reduces it effect.

**Bibliography:**

<https://catalog.data.gov/dataset>

<https://toolbox.google.com/datasetsearch>

<https://www.kdnuggets.com/datasets/index.html>

<https://archive.ics.uci.edu/ml/datasets.php>

<https://www.datacamp.com/?utm_source=adwords_ppc&utm_campaignid=1242944157&utm_adgroupid=58673827368&utm_device=c&utm_keyword=datacamp&utm_matchtype=e&utm_network=g&utm_adpostion=1t1&utm_creative=340731356764&utm_targetid=kwd-297372810188&utm_loc_interest_ms=&utm_loc_physical_ms=20466&gclid=CjwKCAiA_MPuBRB5EiwAHTTvMa1y8NSVn0oGrH_id5PqNxiyDArf_0S6snV5bItuimKCV_VRh1JMjxoCmIwQAvD_BwE>

<https://www.kaggle.com/>