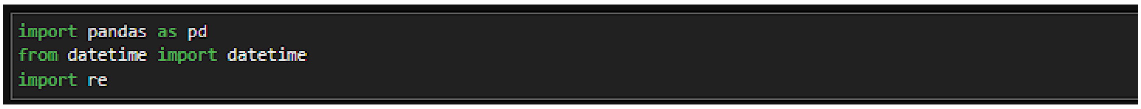
Brief Report of the Work Done

Dataset Used: Adoption of genetically modified crops, percentage of planting of different altered genes over the total plantings. Link: USDA ERS - Adoption of Genetically Engineered Crops in the U.S.

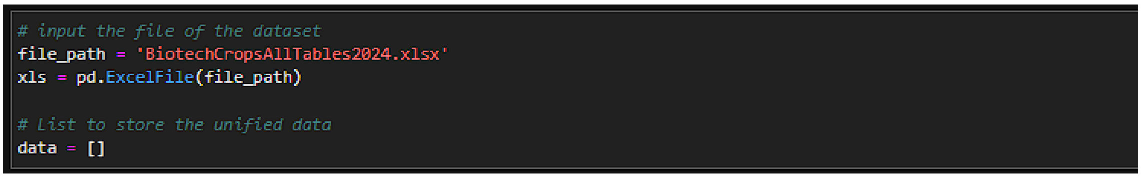
1. Solution Development:

- **Dataset Extraction**: The dataset was extracted directly from the source in .xlsx format, which contained several tables spread across three sheets.
- **Data Modeling**: I used Python and Jupyter Notebooks to process and unify the tables into a single coherent structure. The original file had distributed data, so I designed a custom script to consolidate this information into a single table. Due to the limited volume and structure of the dataset, I opted not to use a Kimball schema for dimension and fact tables.

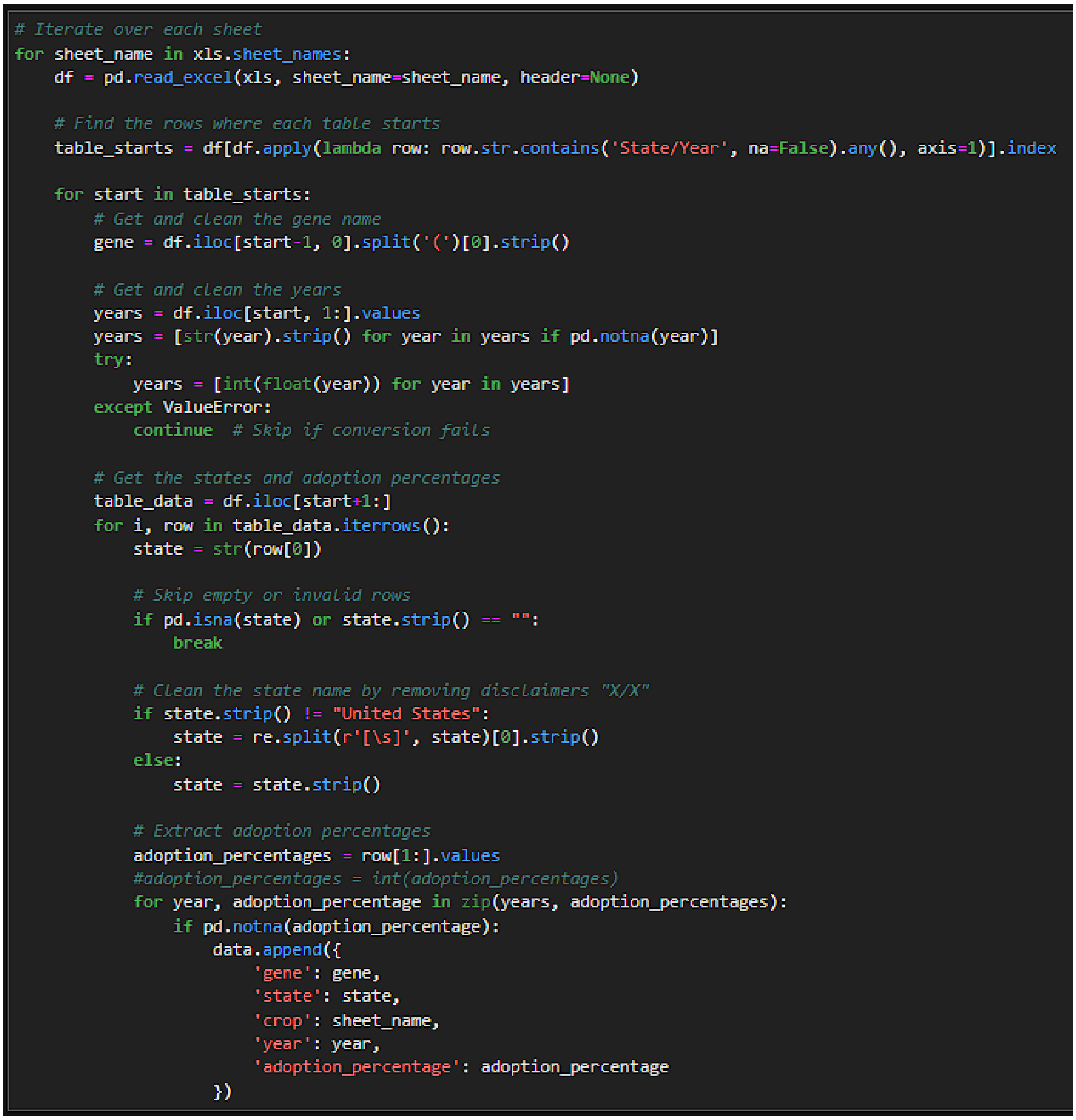
Python Script Used:



```
import pandas as pd
from datetime import datetime
import re
```



```
# input the file of the dataset
file_path = 'BiotechCropsAllTables2024.xlsx'
xls = pd.ExcelFile(file_path)

# List to store the unified data
data = []
```



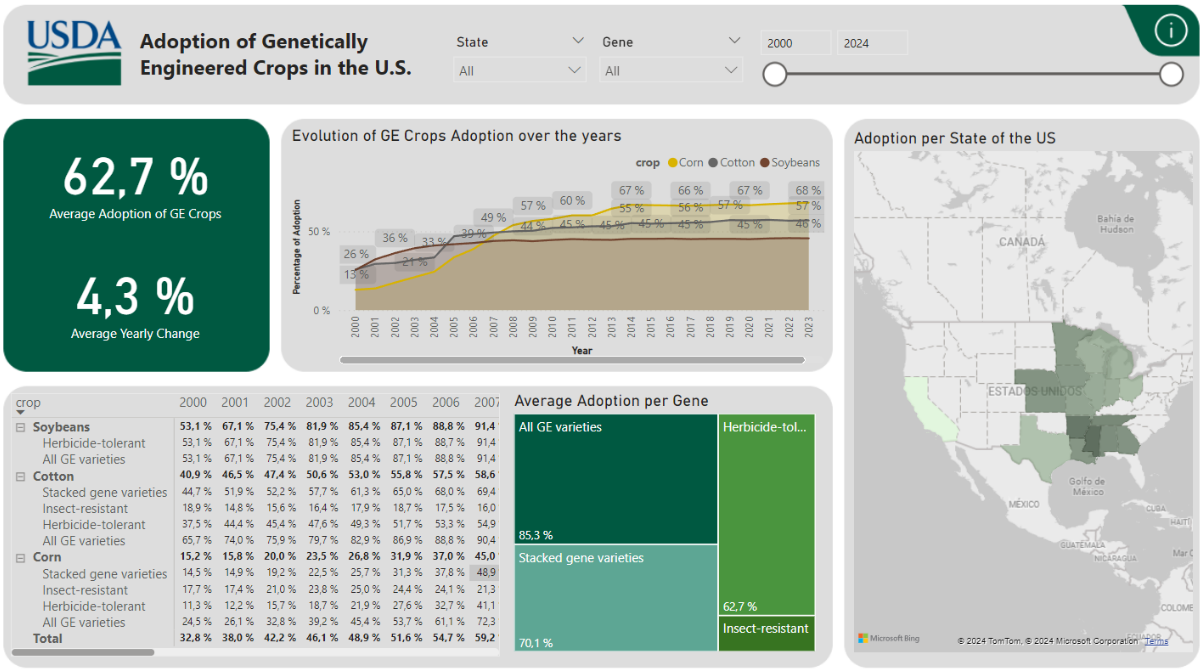
```
# Iterate over each sheet
for sheet_name in xls.sheet_names:
    df = pd.read_excel(xls, sheet_name=sheet_name, header=None)

    # Find the rows where each table starts
    table_starts = df[df.apply(lambda row: row.str.contains('State/Year', na=False).any(), axis=1)].index

    for start in table_starts:
        # Get and clean the gene name
        gene = df.iloc[start-1, 0].split('(')[0].strip()

        # Get and clean the years
        years = df.iloc[start, 1:].values
        years = [str(year).strip() for year in years if pd.notna(year)]
        try:
            years = [int(float(year)) for year in years]
        except ValueError:
            continue  # Skip if conversion fails

        # Get the states and adoption percentages
        table_data = df.iloc[start+1:]
        for i, row in table_data.iterrows():
            state = str(row[0])

            # Skip empty or invalid rows
            if pd.isna(state) or state.strip() == "":
                break

            # Clean the state name by removing disclaimers "X/X"
            if state.strip() != "United States":
                state = re.split(r'[\s]', state)[0].strip()
            else:
                state = state.strip()

            # Extract adoption percentages
            adoption_percentages = row[1:].values
            #adoption_percentages = int(adoption_percentages)
            for year, adoption_percentage in zip(years, adoption_percentages):
                if pd.notna(adoption_percentage):
                    data.append({
                        'gene': gene,
                        'state': state,
                        'crop': sheet_name,
                        'year': year,
                        'adoption_percentage': adoption_percentage
                    })
```

- **Dashboard Design in Power BI**: Importing the modeled data (.xlsx) into Power BI and designing an interactive dashboard with DAX measures such as:
    - Average adoption percentage
    - Average annual growth rate

2. Dashboard Objective

The dashboard aims to provide valuable insights into trends in genetically modified crop adoption over the past two decades by state in the U.S. This is particularly useful for a company in the agricultural sector focused on genetic modification, enabling them to identify:

- **Adoption Trends**: Understanding how the adoption of different genes has evolved over the years and in which states they have been most prevalent.
- **Geographical Information**: Visualizing which regions have the most significant adoption of these technologies, aiding strategic business decisions.



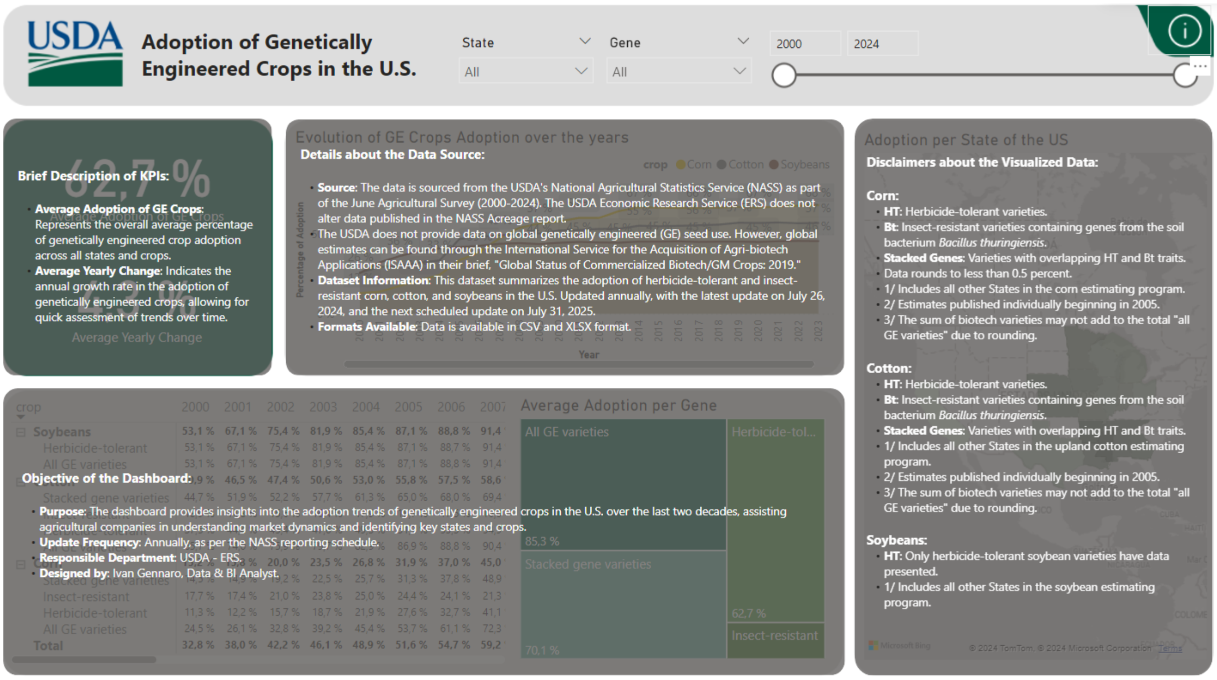
Adoption of Genetically Engineered Crops in the U.S.

State: All | Gene: All | 2000 – 2024

62,7 %
Average Adoption of GE Crops

4,3 %
Average Yearly Change

crop	2000	2001	2002	2003	2004	2005	2006	2007
Soybeans	**53,1 %**	**67,1 %**	**75,4 %**	**81,9 %**	**85,4 %**	**87,1 %**	**88,8 %**	**91,4**
Herbicide-tolerant	53,1 %	67,1 %	75,4 %	81,9 %	85,4 %	87,1 %	88,7 %	91,4
All GE varieties	53,1 %	67,1 %	75,4 %	81,9 %	85,4 %	87,1 %	88,8 %	91,4
Cotton	**40,9 %**	**46,5 %**	**47,4 %**	**50,6 %**	**53,0 %**	**55,8 %**	**57,5 %**	**58,6**
Stacked gene varieties	44,7 %	51,9 %	52,2 %	57,7 %	61,3 %	65,0 %	68,0 %	69,4
Insect-resistant	18,9 %	14,8 %	15,6 %	16,4 %	17,9 %	18,7 %	17,5 %	16,0
Herbicide-tolerant	37,5 %	44,4 %	45,4 %	47,6 %	49,3 %	51,7 %	53,3 %	54,9
All GE varieties	65,7 %	74,0 %	75,9 %	79,7 %	82,9 %	86,9 %	88,8 %	90,4
Corn	**15,2 %**	**15,8 %**	**20,0 %**	**23,5 %**	**26,8 %**	**31,9 %**	**37,0 %**	**45,0**
Stacked gene varieties	14,5 %	14,9 %	19,2 %	22,5 %	25,7 %	31,3 %	37,8 %	48,9
Insect-resistant	17,7 %	17,4 %	21,0 %	23,8 %	25,0 %	24,4 %	24,1 %	21,3
Herbicide-tolerant	11,3 %	12,2 %	15,7 %	18,7 %	21,9 %	27,6 %	32,7 %	41,1
All GE varieties	24,5 %	26,1 %	32,8 %	39,2 %	45,4 %	53,7 %	61,1 %	72,3
Total	**32,8 %**	**38,0 %**	**42,2 %**	**46,1 %**	**48,9 %**	**51,6 %**	**54,7 %**	**59,2**



Adoption of Genetically Engineered Crops in the U.S.

State: All | Gene: All | 2000 – 2024

Brief Description of KPIs:

- **Average Adoption of GE Crops**: Represents the overall average percentage of genetically engineered crop adoption across all states and crops.
- **Average Yearly Change**: Indicates the annual growth rate in the adoption of genetically engineered crops, allowing for quick assessment of trends over time.

Details about the Data Source:

- **Source**: The data is sourced from the USDA's National Agricultural Statistics Service (NASS) as part of the June Agricultural Survey (2000-2024). The USDA Economic Research Service (ERS) does not alter data published in the NASS Acreage report.
- The USDA does not provide data on global genetically engineered (GE) seed use. However, global estimates can be found through the International Service for the Acquisition of Agri-biotech Applications (ISAAA) in their brief, "Global Status of Commercialized Biotech/GM Crops: 2019."
- **Dataset Information**: This dataset summarizes the adoption of herbicide-tolerant and insect-resistant corn, cotton, and soybeans in the U.S. Updated annually, with the latest update on July 26, 2024, and the next scheduled update on July 31, 2025.
- **Formats Available**: Data is available in CSV and XLSX format.

Disclaimers about the Visualized Data:

Corn:

- **HT**: Herbicide-tolerant varieties.
- **Bt**: Insect-resistant varieties containing genes from the soil bacterium *Bacillus thuringiensis*.
- **Stacked Genes**: Varieties with overlapping HT and Bt traits.
- Data rounds to less than 0.5 percent.
- 1/ Includes all other States in the corn estimating program.
- 2/ Estimates published individually beginning in 2005.
- 3/ The sum of biotech varieties may not add to the total "all GE varieties" due to rounding.

Cotton:

- **HT**: Herbicide-tolerant varieties.
- **Bt**: Insect-resistant varieties containing genes from the soil bacterium *Bacillus thuringiensis*.
- **Stacked Genes**: Varieties with overlapping HT and Bt traits.
- 1/ Includes all other States in the upland cotton estimating program.
- 2/ Estimates published individually beginning in 2005.
- 3/ The sum of biotech varieties may not add to the total "all GE varieties" due to rounding.

Soybeans:

- **HT**: Only herbicide-tolerant soybean varieties have data presented.
- 1/ Includes all other States in the soybean estimating program.

Objective of the Dashboard:

- **Purpose**: The dashboard provides insights into the adoption trends of genetically engineered crops in the U.S. over the last two decades, assisting agricultural companies in understanding market dynamics and identifying key states and crops.
- **Update Frequency**: Annually, as per the NASS reporting schedule.
- **Responsible Department**: USDA - ERS.
- **Designed by**: Ivan Gennaro, Data & BI Analyst.

3. Conclusion

Given the problem presented, I opted for an efficient solution planned over three days:

- **Day 1**: Dataset search, BI planning, and data modeling development.
- **Day 2**: Preliminary dashboard design and data importation.
- **Day 3**: Visualization refinement, creation of DAX measures, and documentation.

Total time invested: Approximately 5 hours.