Portfolio

Name: Amey Arun Parange
Title: Business Intelligence Developer
Email id: parange.a@northeastern.edu
LinkedIn: https://www.linkedin.com/in/amey-parange
Github: amey379/Portfolio (github.com)

Summary:
Hello! I am an experienced Business Intelligence professional with a strong background in data warehousing, dimensional modeling, ETL development, and data visualization. I have over three years of experience working with the fintech team at Priceline.com, where I implemented and maintained ETL pipelines using Informatica PowerCenter and PySpark. My work involved integrating data into Oracle ERP-based accounting systems and a Finance Data Warehouse, ensuring accurate and timely data delivery.

I am skilled in developing dashboards and reports that facilitate advanced decision-making, using tools such as SQL and Tableau. This portfolio showcases my Business Intelligence projects during my Master's in Information Systems, as well as my experience as a Business Intelligence Teaching Assistant, where I mentored 90 students. It highlights my proficiency in handling large datasets across various domains, transforming them into meaningful insights that empower stakeholders to make informed decisions.

Skills Highlighted:

Coding Languages: Python, SQL, Oracle PL/SQL, PySpark, Java, Unix Shell Script

Databases: Oracle, MS SQL Server, Snowflake, Cosmos, AWS S3, MySQL

Cloud and Big Data: Google Cloud Platform, Microsoft Azure, Dataproc, Spark, Hadoop,

Data Engineering Tools: Informatica, Talend, Azure Data Factory, SSIS, Power Platform, Databricks

BI and Other Tools: Power BI, Tableau, ER Studio, Looker, Git, Jira, Excel, MS Power Automate, Alteryx

Python Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Beautiful Soup

Table of Contents

IMDB Movie Analysis

Problem Statement

The objective of this project is to develop an analytical solution for raw movie data, encompassing aspects such as crew, box office performance, movie details, and name changes. The goal is to implement a data warehousing solution and create impactful dashboards that address various business requirements related to IMDB Movies.

Source Data:

- ***MySQL*** : Load this SQL DB file to your local MySQL database and use as source connection
- ***Revenue related TSV files*** with Movie box office numbers
- ***SCD2 related data JSON files***: There are 2 JSON file contains Movie titles changes file contains actor name changes

BI Requirements:

Trend Analysis, Genre Analysis, Performance Metrics, Director Success Metrics, Actor and Actress Film Records, Seasonal Analysis, Release Regions and Movie Success.

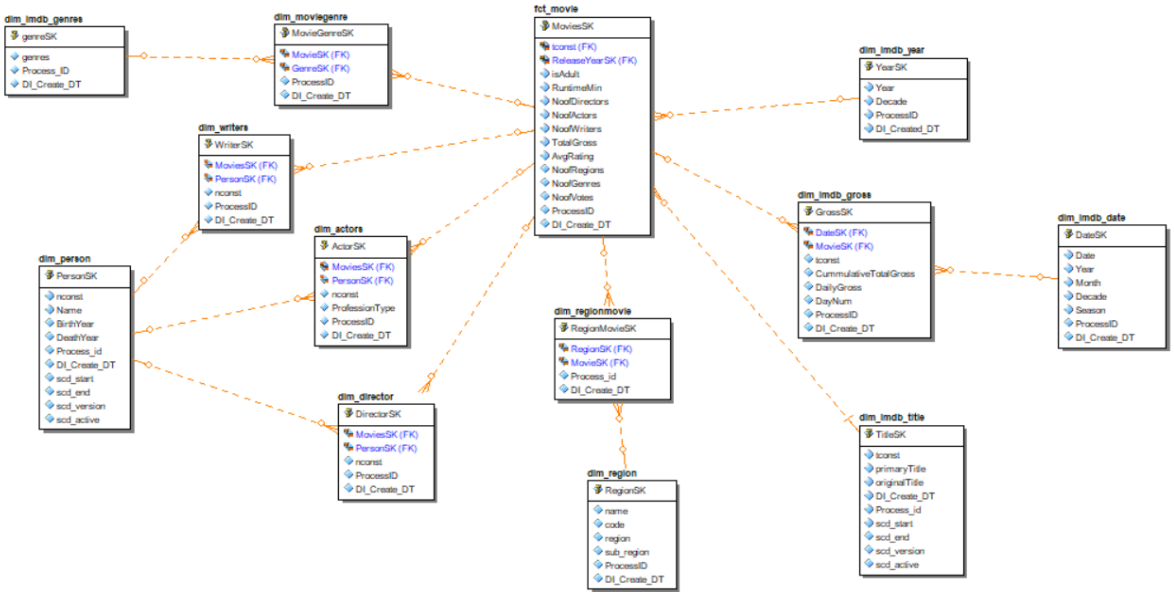
Skills:

Python, SQL, Alteryx, ER Studio, Talend, Power BI, Tableau

Solution:

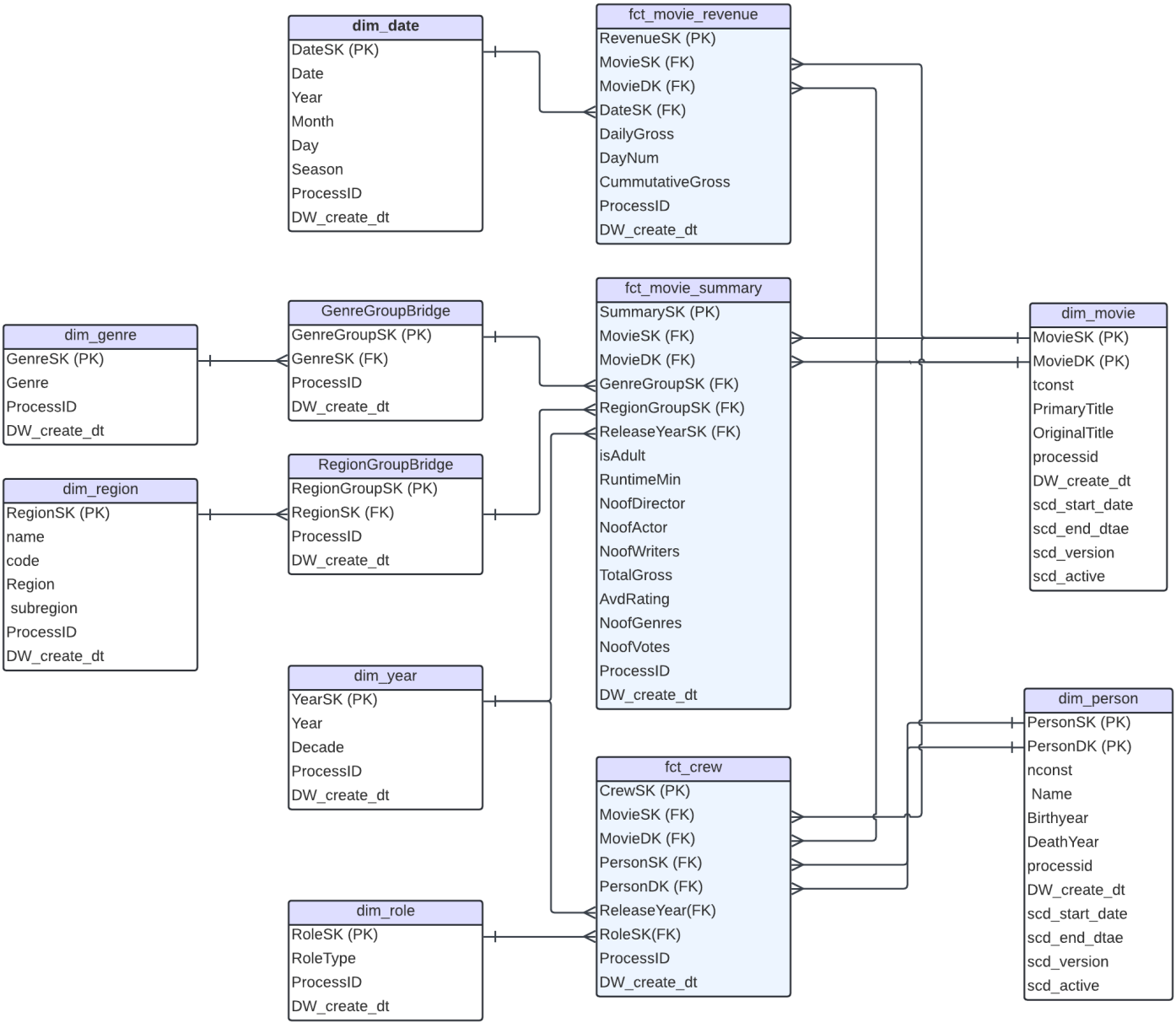
Github: [Portfolio/IMDBMoviesDataWarehousing at main · amey379/Portfolio (github.com)]

1. **Data Profiling:** Understand data structure and quality. Developed workflow in Alteryx to profile data. Documentation Available on github.
2. **Dimensional Modeling:**
   I created this Data Model in Dec 2023 for the Project. I have updated this model after this.



Updated Model:

1. Reduced dependency
2. Implement Conformed Dimensions
3. Keys Consistency
4. Improved Performance



3. **Data Staging:**
   Staged data from the source as it is.
   Formatted data according to profiling results eg. Formatted date and implemented data types.

4. **Data Curation:**
   In this step, I cleaned, **normalized** and standardized data to prepare it according to business requirements. Prepared validation rules for every table and rejected rows into a temp table for evaluation.

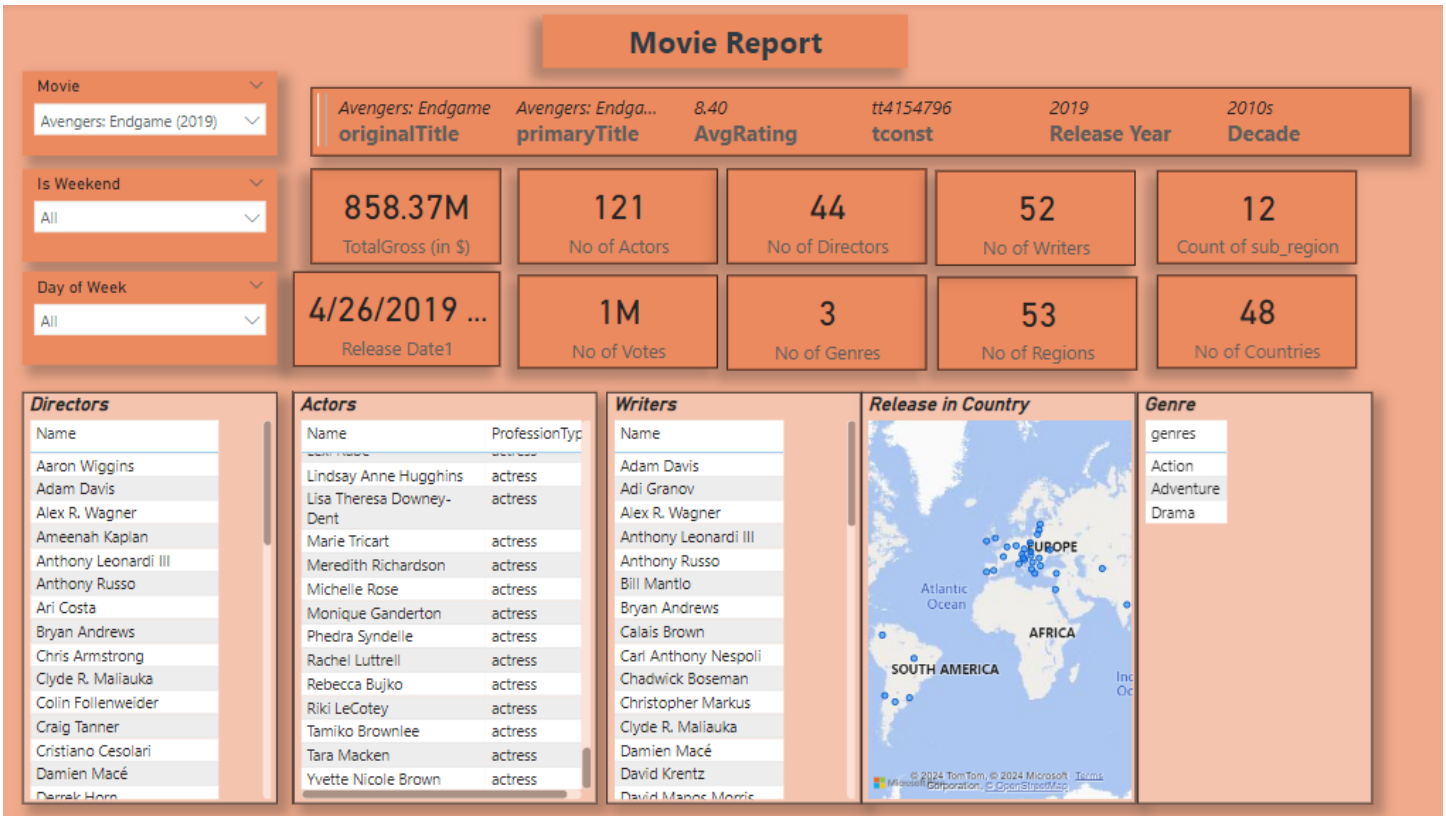
5. **Data Integration**
   In this step, data is integrated into landing tables from multiple tables required for loading facts and dimensions.

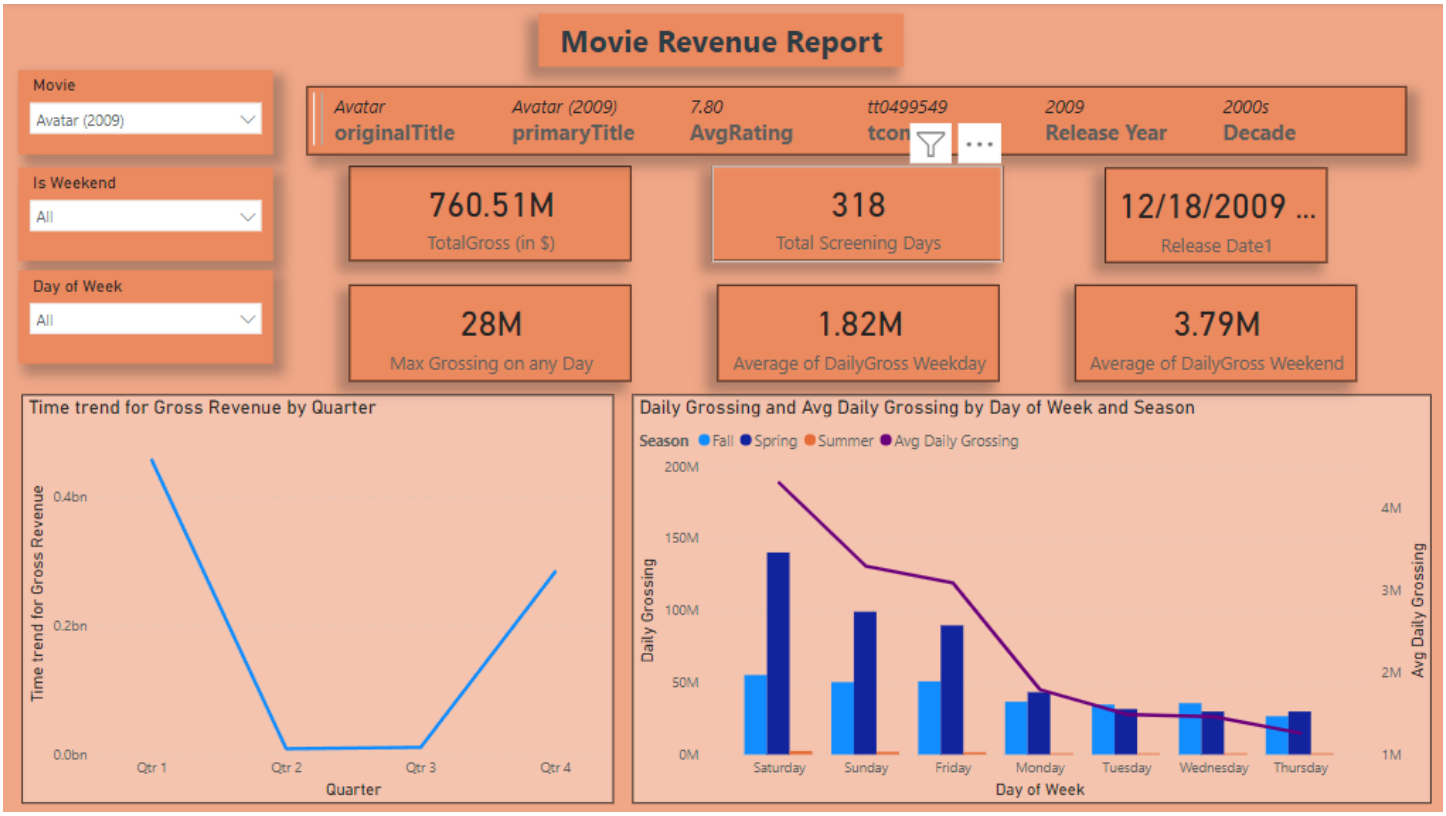
6. **Load Facts and Dimensions**
   Implemented workflow to load data into a dimensional model. As some actors changed names over time, I implemented Slowly changing dimensions Type 2.

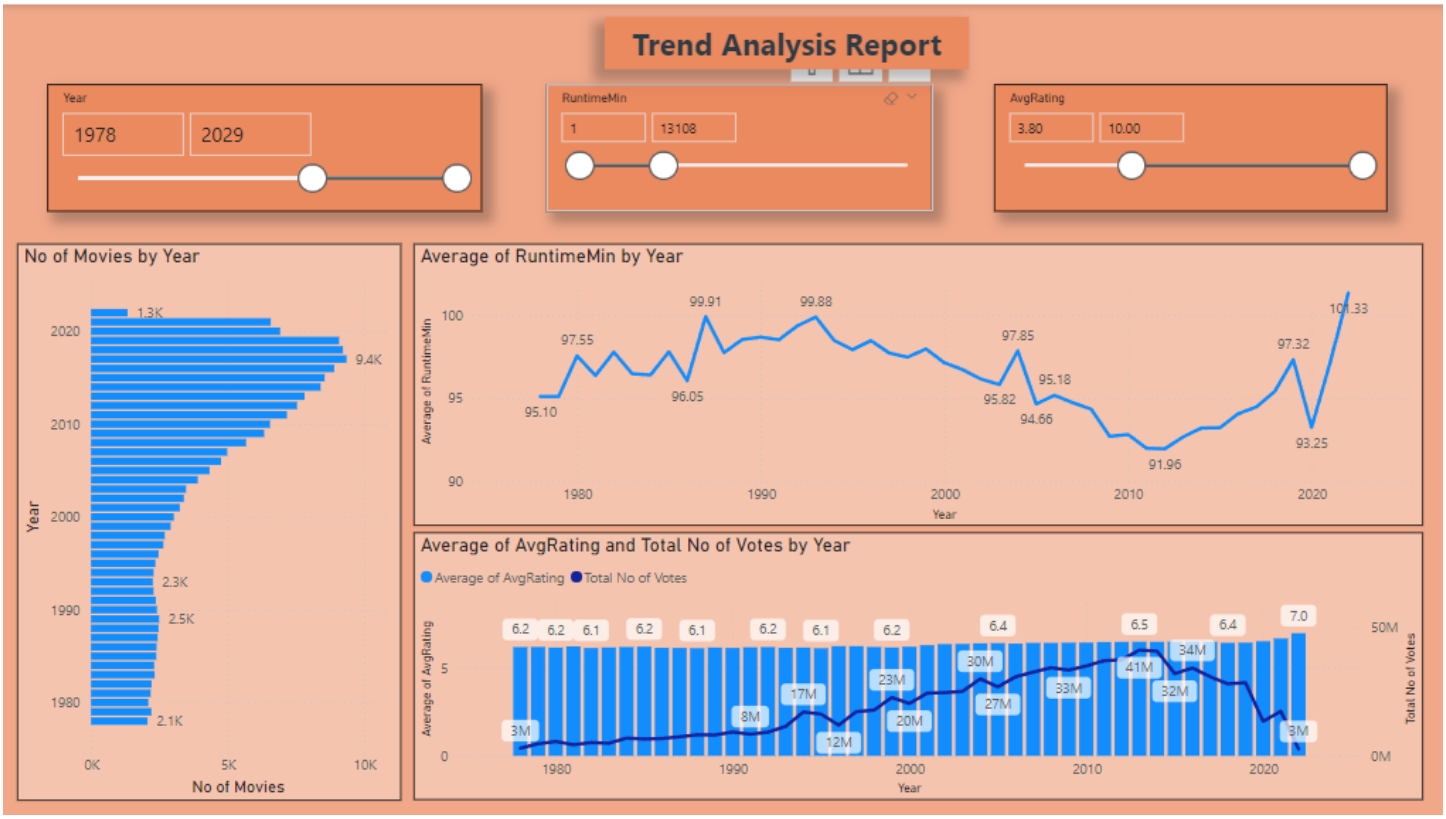
7. **Data Visualization**

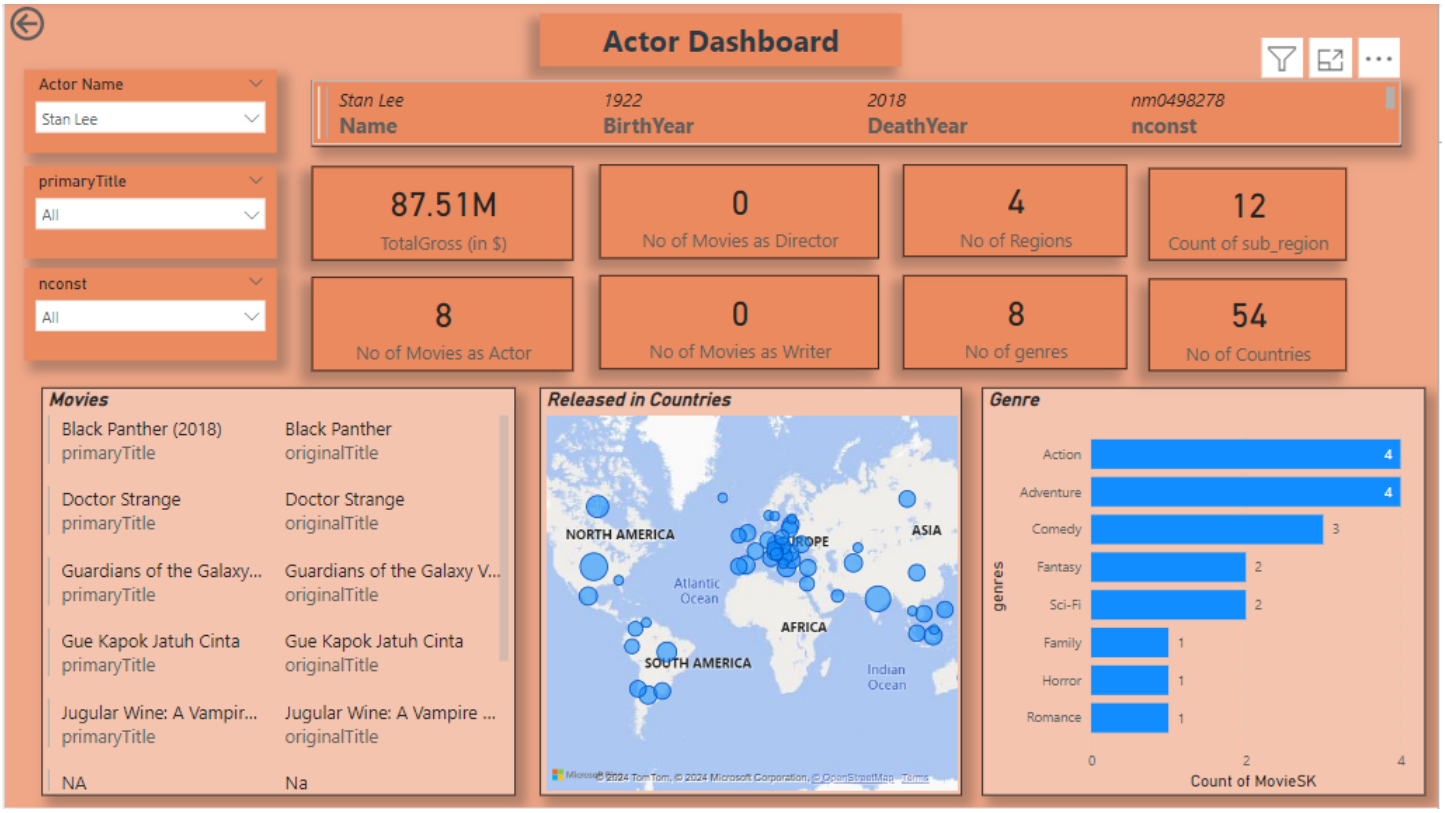
BI Reports

Reports are based on the old Dimensional Model. I have used DAX and Power Query to enhance my reports. I have implemented similar dashboards in Tableau. (Available on Github)

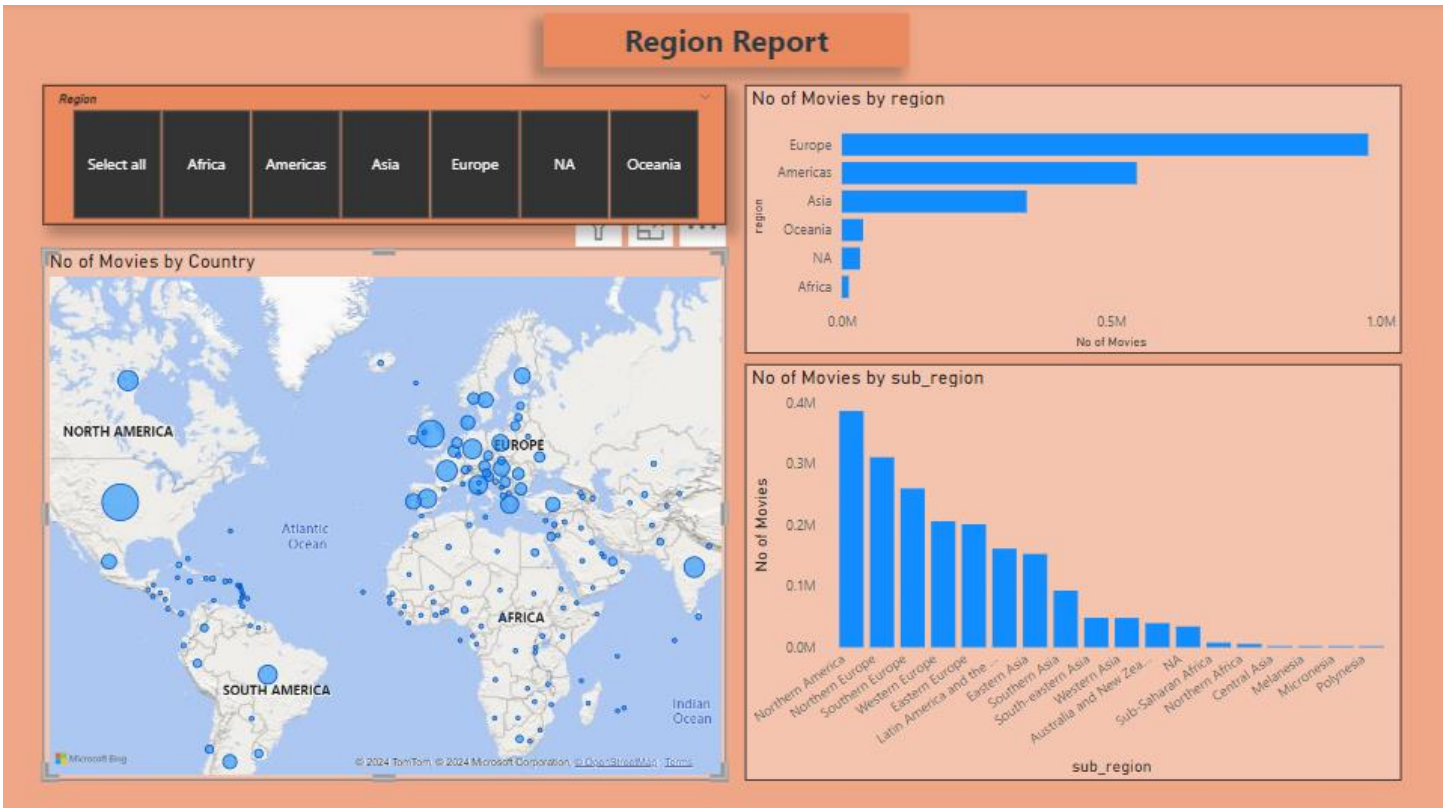



Movie Revenue Report
Movie
Avatar (2009)
Is Weekend
All
Day of Week
All
Avatar
originalTitle
Avatar (2009)
primaryTitle
7.80
AvgRating
tt0499549
tcon
2009
Release Year
2000s
Decade
760.51M
TotalGross (in $)
318
Total Screening Days
12/18/2009 ...
Release Date1
28M
Max Grossing on any Day
1.82M
Average of DailyGross Weekday
3.79M
Average of DailyGross Weekend
Time trend for Gross Revenue by Quarter
Time trend for Gross Revenue
0.4bn
0.2bn
0.0bn
Qtr 1
Qtr 2
Qtr 3
Qtr 4
Quarter
Daily Grossing and Avg Daily Grossing by Day of Week and Season
Season
Fall
Spring
Summer
Avg Daily Grossing
Daily Grossing
200M
150M
100M
50M
0M
4M
3M
2M
1M
Avg Daily Grossing
Saturday
Sunday
Friday
Monday
Tuesday
Wednesday
Thursday
Day of Week


Trend Analysis Report
Year
1978
2029
RuntimeMin
1
13108
AvgRating
3.80
10.00
No of Movies by Year
2020
2010
2000
1990
1980
Year
1.3K
9.4K
2.3K
2.5K
2.1K
0K
5K
10K
No of Movies
Average of RuntimeMin by Year
Average of RuntimeMin
100
95
90
95.10
97.55
96.05
99.91
99.88
97.85
95.82
95.18
94.66
91.96
97.32
93.25
101.33
1980
1990
2000
2010
2020
Year
Average of AvgRating and Total No of Votes by Year
Average of AvgRating
Total No of Votes
Average of AvgRating
5
0
6.2
6.2
6.1
6.2
6.1
6.2
6.1
6.2
6.4
6.5
6.4
7.0
3M
8M
17M
12M
23M
20M
30M
27M
33M
41M
32M
34M
3M
50M
0M
Total No of Votes
1980
1990
2000
2010
2020
Year


Actor Dashboard
Actor Name
Stan Lee
primaryTitle
All
nconst
All
Stan Lee
Name
1922
BirthYear
2018
DeathYear
nm0498278
nconst
87.51M
TotalGross (in $)
0
No of Movies as Director
4
No of Regions
12
Count of sub_region
8
No of Movies as Actor
0
No of Movies as Writer
8
No of genres
54
No of Countries
Movies
Black Panther (2018)
primaryTitle
Black Panther
originalTitle
Doctor Strange
primaryTitle
Doctor Strange
originalTitle
Guardians of the Galaxy...
primaryTitle
Guardians of the Galaxy V...
originalTitle
Gue Kapok Jatuh Cinta
primaryTitle
Gue Kapok Jatuh Cinta
originalTitle
Jugular Wine: A Vampir...
primaryTitle
Jugular Wine: A Vampire ...
originalTitle
NA
Na
Released in Countries
NORTH AMERICA
EUROPE
ASIA
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
Genre
Action
Adventure
Comedy
Fantasy
Sci-Fi
Family
Horror
Romance
genres
Count of MovieSK


Director Dashboard
Director Name
A.J. Briones
primaryTitle
All
nconst
All
A.J. Briones
Name
9999
BirthYear
9999
DeathYear
nm2310077
nconst
152.10M
TotalGross (in $)
5
No of Movies as Director
4
No of Regions
12
Count of sub_region
0
No of Movies as Actor
4
No of Movies as Writer1
7
No of genres
56
No of Countries
Movies
A Night of Horror: Nigh...
primaryTitle
A Night of Horror: Nightm...
originalTitle
Avatar (2009)
primaryTitle
Avatar
originalTitle
Battleship
primaryTitle
Battleship
originalTitle
Dawn of the Planet of t...
primaryTitle
Dawn of the Planet of the ...
originalTitle
NA
primaryTitle
Na
originalTitle
Movie released countries
NORTH AMERICA
EUROPE
ASIA
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
Genre
Action
Adventure
Drama
Fantasy
Horror
Mystery
Sci-Fi
genres
Count of MovieSK


Region Report
Region
Select all
Africa
Americas
Asia
Europe
NA
Oceania
No of Movies by Country
NORTH AMERICA
EUROPE
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
No of Movies by region
Europe
Americas
Asia
Oceania
NA
Africa
region
0.0M
0.5M
1.0M
No of Movies
No of Movies by sub_region
0.4M
0.3M
0.2M
0.1M
0.0M
No of Movies
Northern America
Northern Europe
Southern Europe
Western Europe
Eastern Europe
Latin America and the ...
Eastern Asia
Southern Asia
South-eastern Asia
Western Asia
Australia and New Zea...
NA
Sub-Saharan Africa
Northern Africa
Central Asia
Melanesia
Micronesia
Polynesia
sub_region

Food Facilities Inspection

Problem Statement

The objective of this project is to develop a report for Food Facilities Inspection. The goal is to implement a data warehousing solution and create impactful dashboards that address various business requirements related to Inspections and Violations.

Source Data: Public Dataset, [Food Inspections | City of Chicago | Data Portal]

BI Requirements:

Monthly Report, Facility Report, Inspection Report, Worst 10 Businesses and Regional Dashboards.

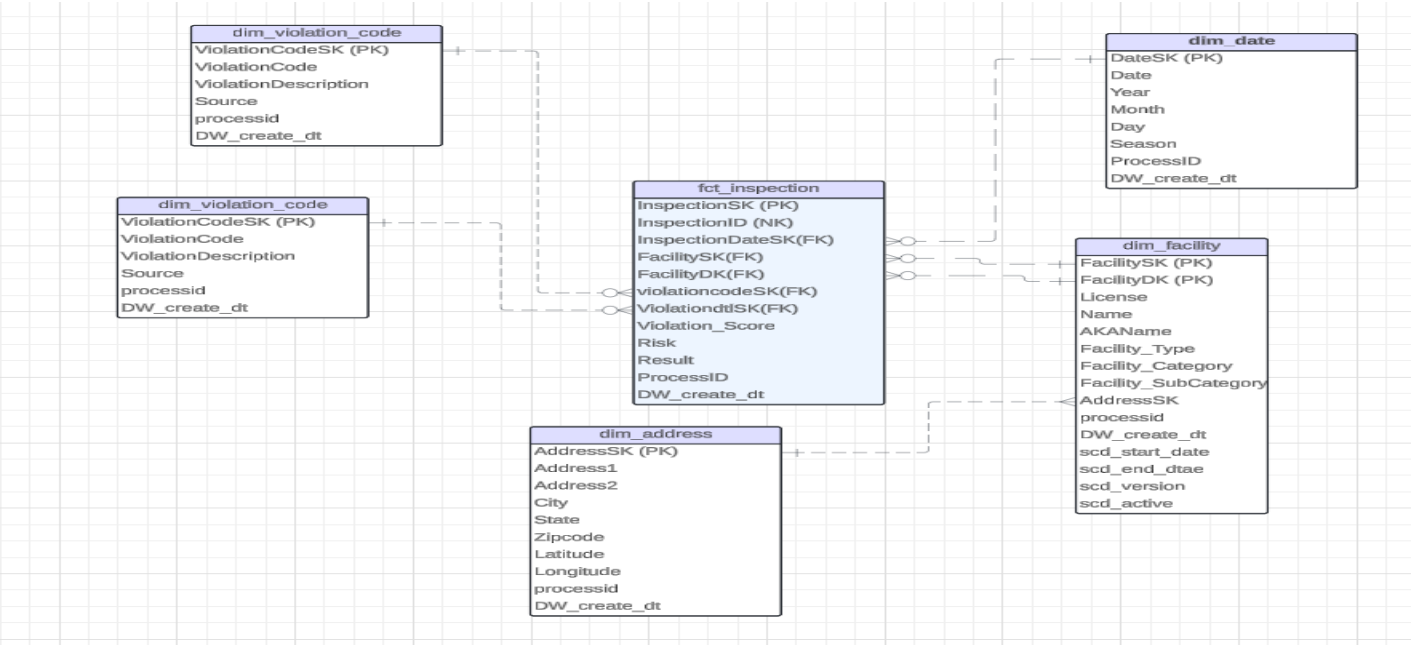
Skills:

Azure Data Factory, Azure Data Lake Storage, Azure SQL Database, Power BI, Tableau, SQL, DAX

Solution:

Github: [Portfolio/FoodFacilitiesInspectionAnalytics/README.md at main · amey379/Portfolio (github.com)]

1. **Data Profiling:** Understand data structure and quality. Developed workflow in Alteryx to profile data. Documentation Available on github.
2. **Dimensional Modeling:** Developed a star schema data model



3. **Data Staging:**
   Staged data from the source as it is into Azure Data Lake Storage.

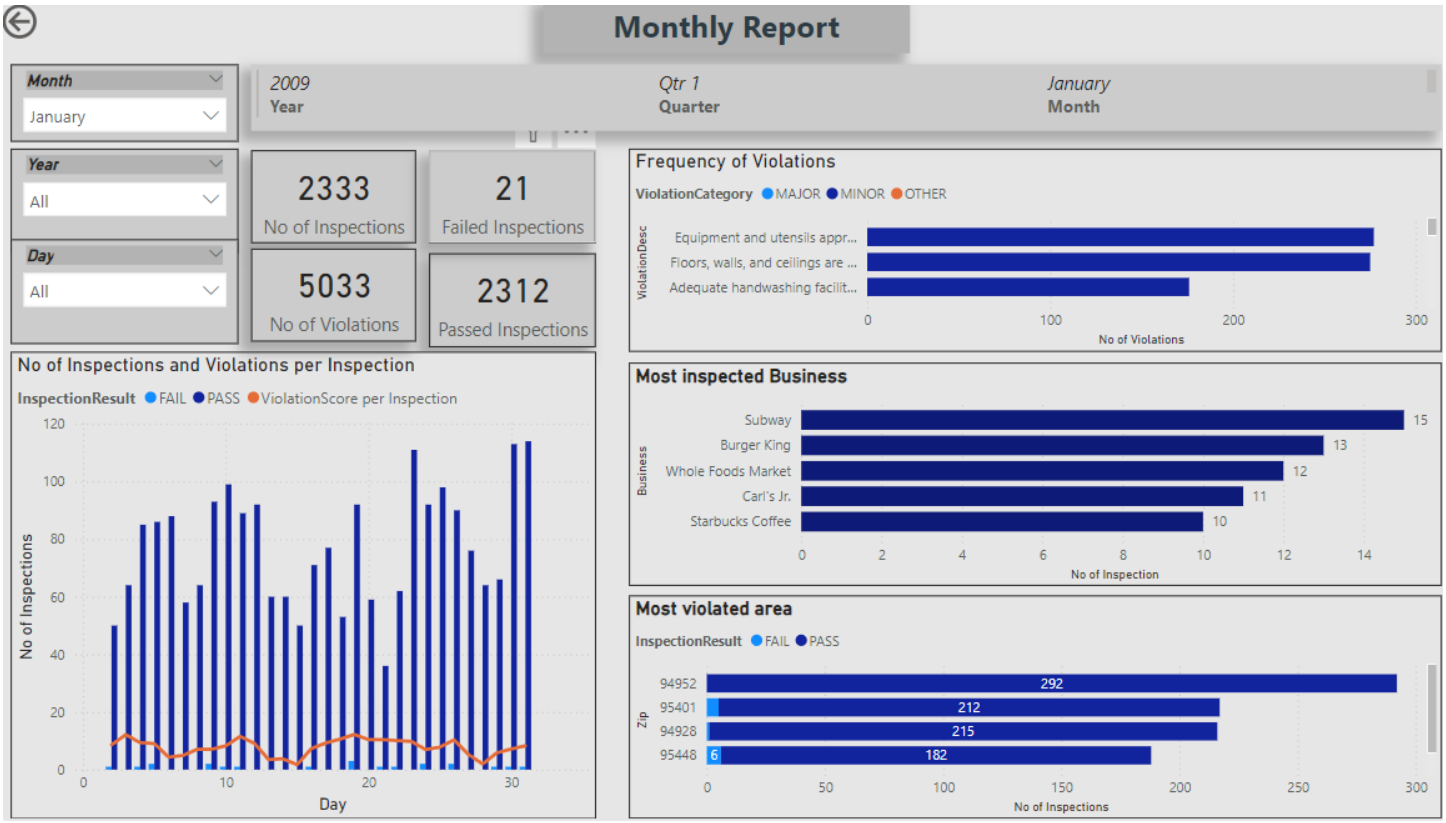
4. **Data Curation:**
   In this step, I cleaned, **normalized** and standardized data to prepare it according to business requirements. Prepared validation rules for every table and rejected rows into a temp file for evaluation. Curated data is storage in Azure Data Lake Storage

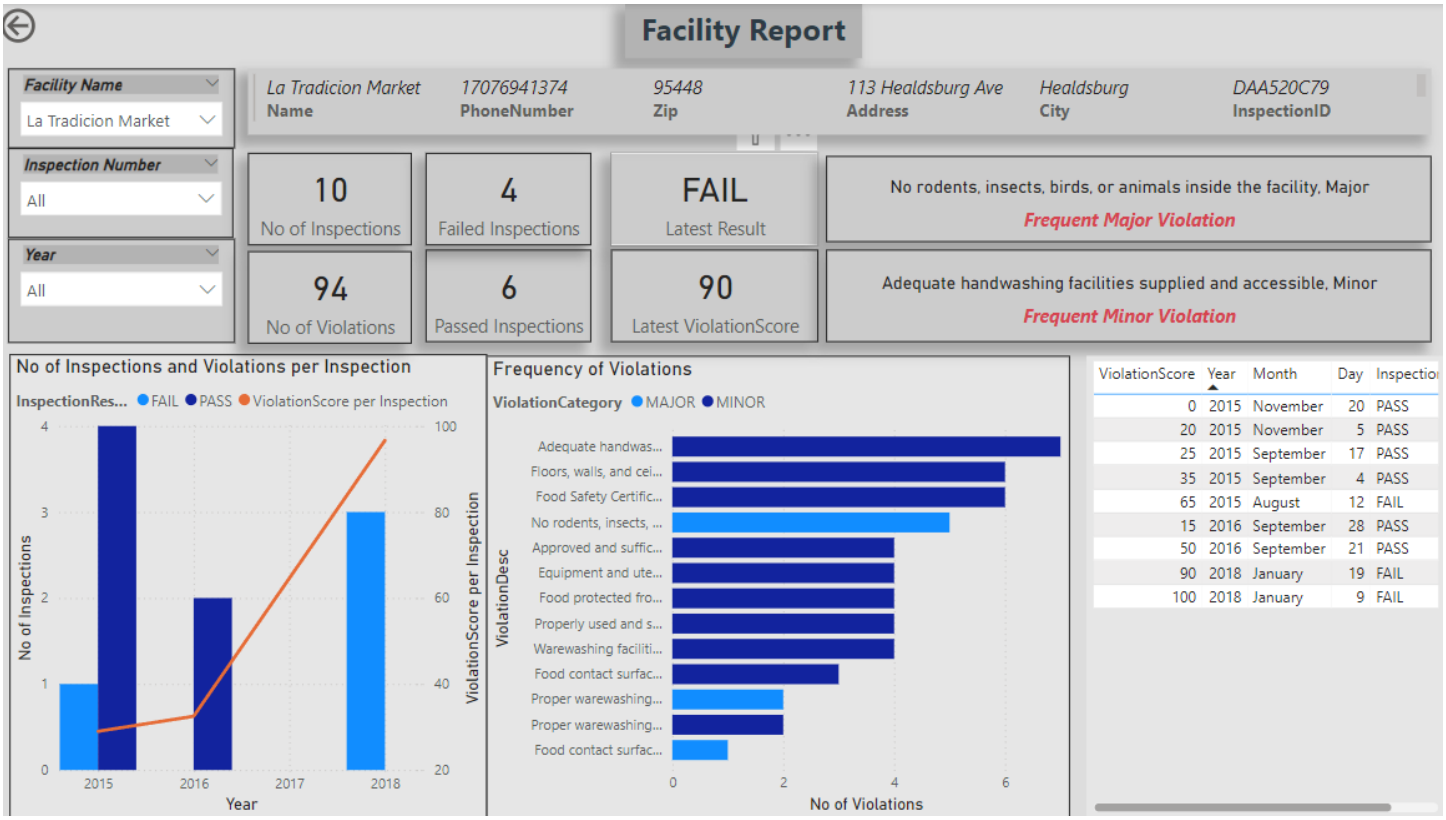
5. **Data Integration**
   In this step, data is integrated into landing tables from multiple tables required for loading facts and dimensions. The landing table was maintained in Azure SQL Database.

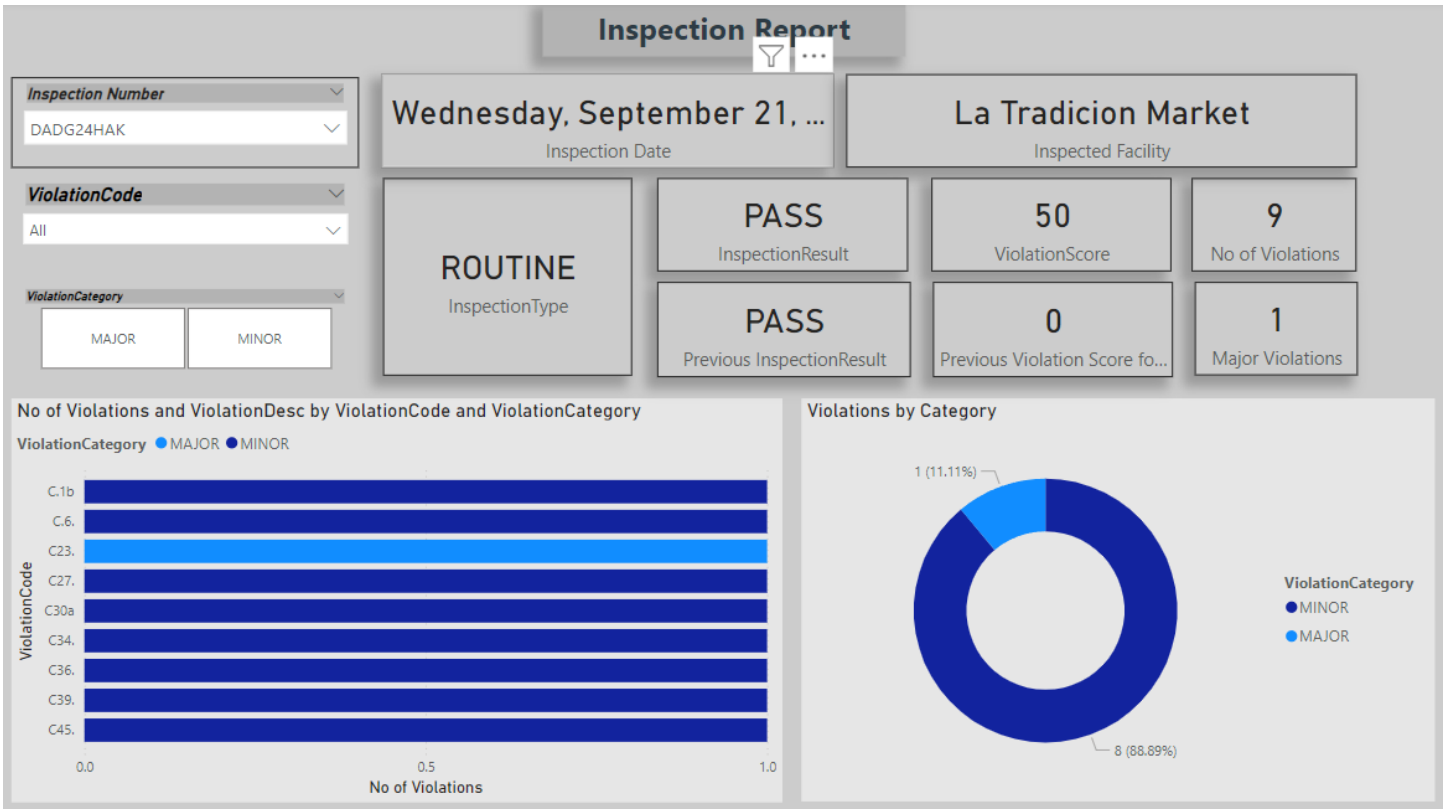
6. **Load Facts and Dimensions**
   Implemented workflow to load data into a dimensional model.

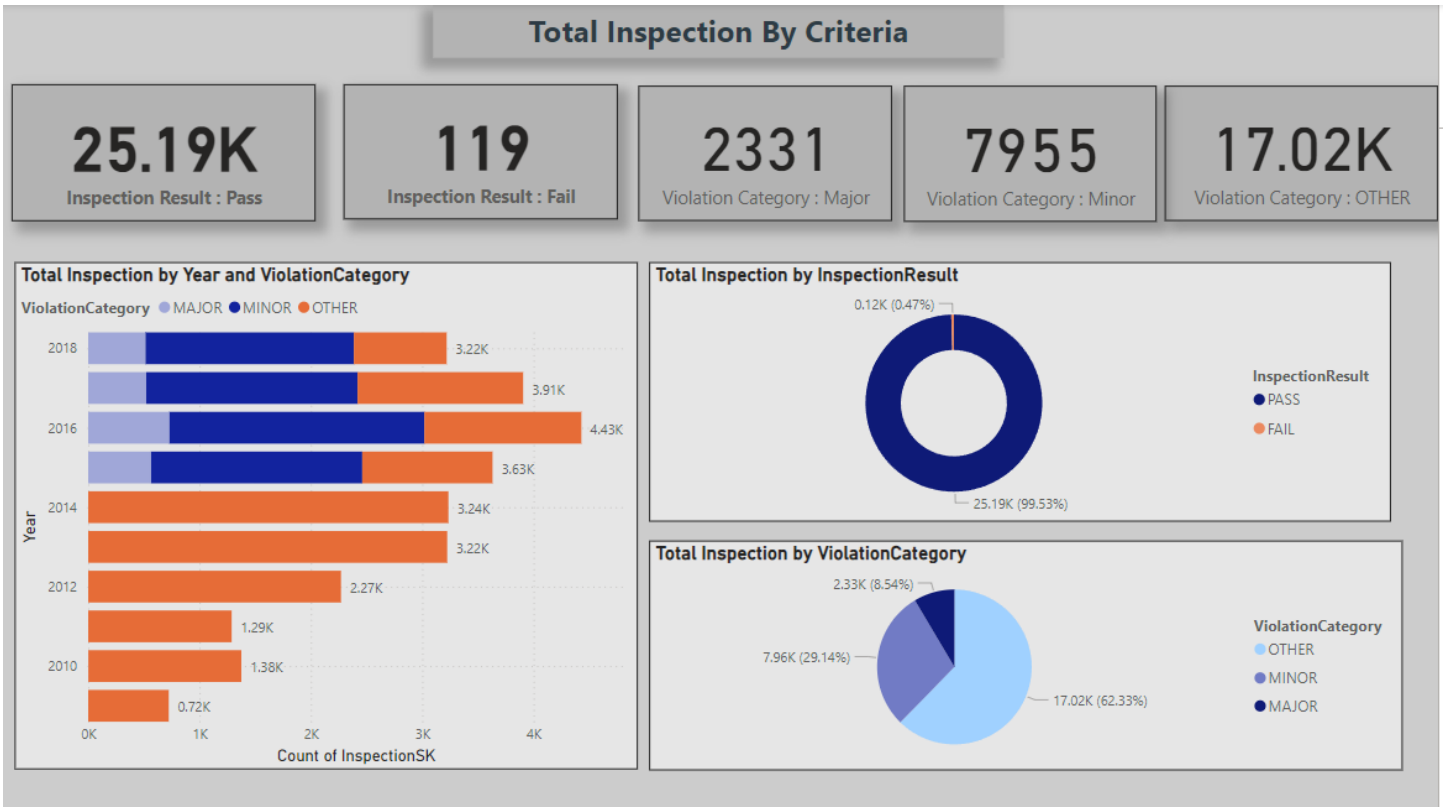
7. **Data Visualization**

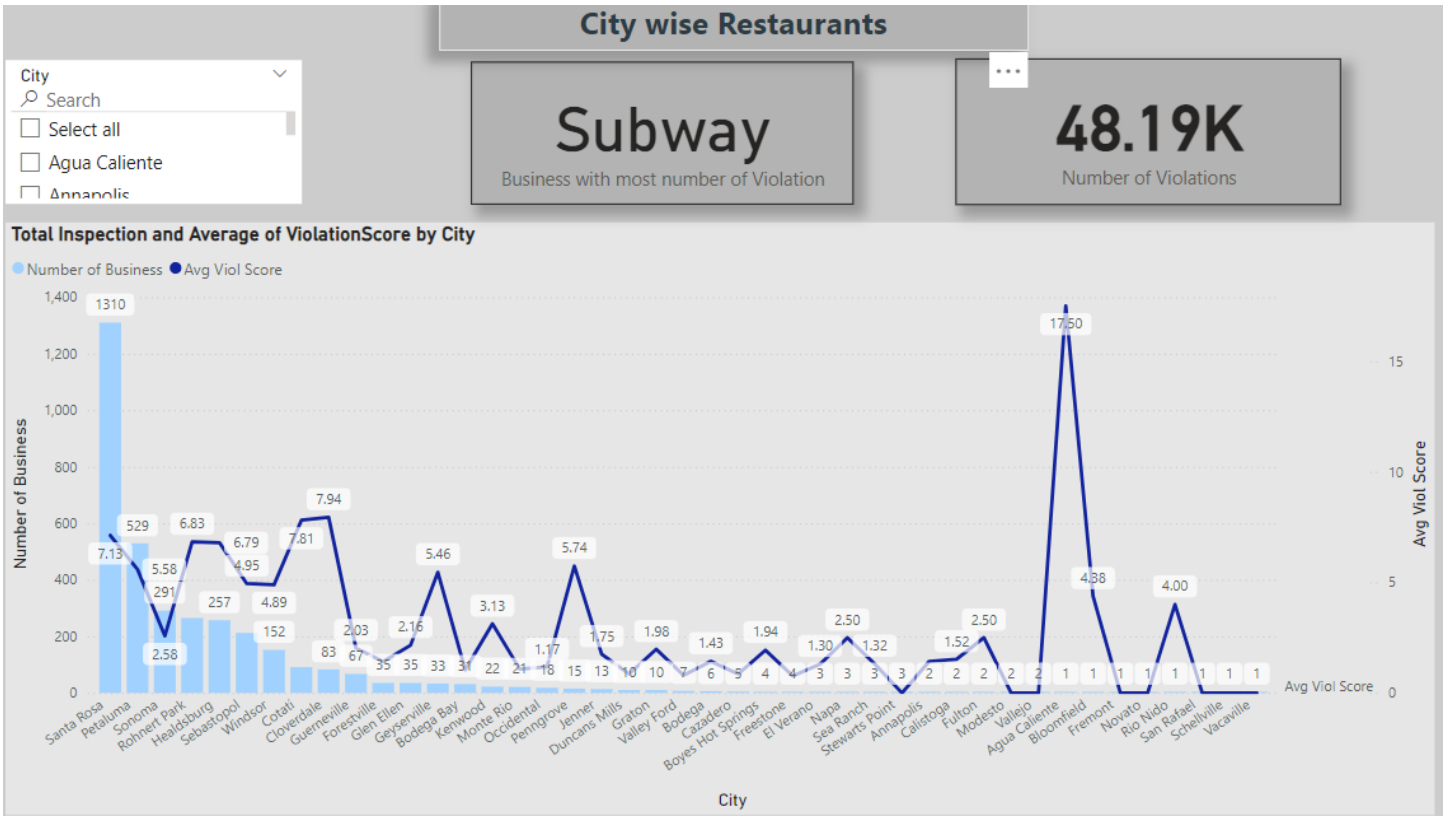
Inspection Dashboards




Facility Report
Facility Name
La Tradicion Market
Inspection Number
All
Year
All
La Tradicion Market
Name
17076941374
PhoneNumber
95448
Zip
113 Healdsburg Ave
Address
Healdsburg
City
DAA520C79
InspectionID
10
No of Inspections
4
Failed Inspections
FAIL
Latest Result
No rodents, insects, birds, or animals inside the facility, Major
Frequent Major Violation
94
No of Violations
6
Passed Inspections
90
Latest ViolationScore
Adequate handwashing facilities supplied and accessible, Minor
Frequent Minor Violation
No of Inspections and Violations per Inspection
InspectionRes... FAIL PASS ViolationScore per Inspection
No of Inspections
ViolationScore per Inspection
Year
2015
2016
2017
2018
Frequency of Violations
ViolationCategory MAJOR MINOR
ViolationDesc
Adequate handwas...
Floors, walls, and cei...
Food Safety Certific...
No rodents, insects, ...
Approved and suffic...
Equipment and ute...
Food protected fro...
Properly used and s...
Warewashing faciliti...
Food contact surfac...
Proper warewashing...
Proper warewashing...
Food contact surfac...
No of Violations
ViolationScore Year Month Day Inspectio
0 2015 November 20 PASS
20 2015 November 5 PASS
25 2015 September 17 PASS
35 2015 September 4 PASS
65 2015 August 12 FAIL
15 2016 September 28 PASS
50 2016 September 21 PASS
90 2018 January 19 FAIL
100 2018 January 9 FAIL


Inspection Report
Inspection Number
DADG24HAK
ViolationCode
All
ViolationCategory
MAJOR
MINOR
Wednesday, September 21, ...
Inspection Date
La Tradicion Market
Inspected Facility
ROUTINE
InspectionType
PASS
InspectionResult
50
ViolationScore
9
No of Violations
PASS
Previous InspectionResult
0
Previous Violation Score fo...
1
Major Violations
No of Violations and ViolationDesc by ViolationCode and ViolationCategory
ViolationCategory MAJOR MINOR
ViolationCode
C.1b
C.6.
C23.
C27.
C30a
C34.
C36.
C39.
C45.
No of Violations
Violations by Category
1 (11.11%)
8 (88.89%)
ViolationCategory
MINOR
MAJOR


Total Inspection By Criteria
25.19K
Inspection Result : Pass
119
Inspection Result : Fail
2331
Violation Category : Major
7955
Violation Category : Minor
17.02K
Violation Category : OTHER
Total Inspection by Year and ViolationCategory
ViolationCategory MAJOR MINOR OTHER
Year
2018
2016
2014
2012
2010
3.22K
3.91K
4.43K
3.63K
3.24K
3.22K
2.27K
1.29K
1.38K
0.72K
0K
1K
2K
3K
4K
Count of InspectionSK
Total Inspection by InspectionResult
0.12K (0.47%)
25.19K (99.53%)
InspectionResult
PASS
FAIL
Total Inspection by ViolationCategory
2.33K (8.54%)
7.96K (29.14%)
17.02K (62.33%)
ViolationCategory
OTHER
MINOR
MAJOR


City wise Restaurants
City
Search
Select all
Agua Caliente
Annapolis
Subway
Business with most number of Violation
48.19K
Number of Violations
Total Inspection and Average of ViolationScore by City
Number of Business
Avg Viol Score
Number of Business
Avg Viol Score
City
Santa Rosa
Petaluma
Sonoma
Rohnert Park
Healdsburg
Sebastopol
Windsor
Cotati
Cloverdale
Guerneville
Forestville
Glen Ellen
Geyserville
Bodega Bay
Kenwood
Monte Rio
Occidental
Penngrove
Jenner
Duncans Mills
Graton
Valley Ford
Bodega
Cazadero
Boyes Hot Springs
Freestone
El Verano
Napa
Sea Ranch
Stewarts Point
Annapolis
Calistoga
Fulton
Modesto
Vallejo
Agua Caliente
Bloomfield
Fremont
Novato
Rio Nido
San Rafael
Schellville
Vacaville