Student performance

BI Project Report

Presented By:

-Farah Chihi

-Mohamed Amine Hamadi

-Wadii Selmane

introduction :

Tunis Business School has asked us to help it gain insights about its student’s academic performance accross different dimensions for the past 3 years .we have adopted different processes in order to integrate and use the data that is useful in responding to the school’s objectives .

the administration has asked for an answer for the following questions:

-how does the student performance vary across genders?
-how does the student performance vary across different (freshman,sophomore,Majors)levels?
-how does the student performance fluctuate in relation to state of origin of the alumnis
-how does the average GPA differ through the years?

-what is the correlation between the students' highschool branch and their academic results?
-to which major do the top performing students belong to?

- ***the main phases implemented in the database are:***

1/data generation:

In the creation of the university databases, our team employed an AI tool to generate datasets encompassing three distinct categories: Freshman students represented in JSON format, Sophomore students in Excel, and students with designated majors stored in CSV files.

2/ETL Process(using python):

The transformation procedures comprised a series of intricate steps aimed at refining the dataset:

a)Duplicate Data Removal:identifying and eliminating instances of duplicated data, ensuring the integrity and accuracy of the dataset.

b)Null Value Elimination:identifying null values and eliminating them in order to enhance the dataset's completeness

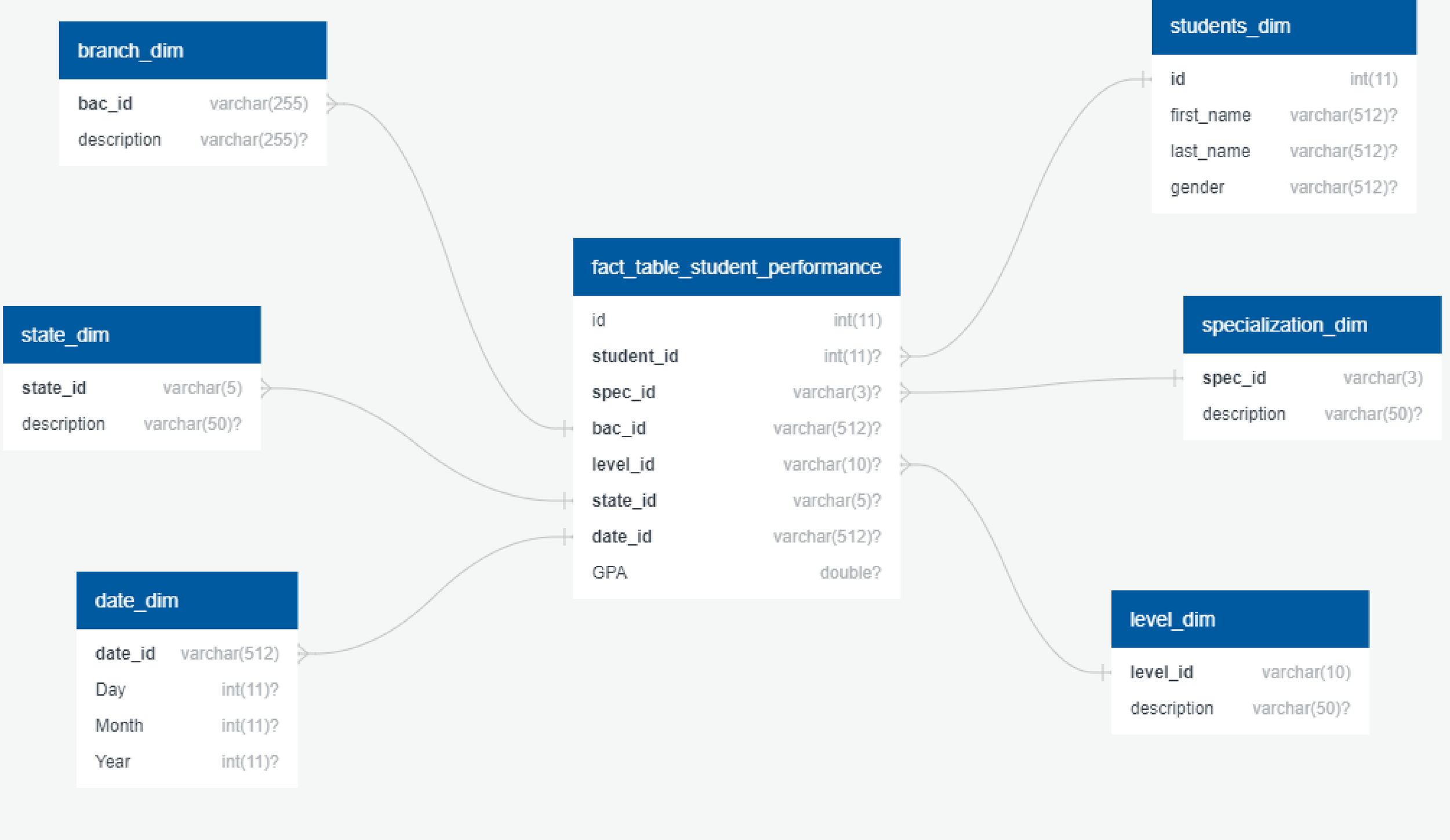
c)unecessary column removal: identifying and eliminating columns that did not contribute to the overall analysis in order to improve efficiency of the data set

d)column merging:two relevant columns were merged to consolidate related information, resulting in a more coherent and consolidated dataset structure.

e)Column Format Modification: To enhance the usability of the data, specific columns underwent modifications in their formats, ensuring a standardized and user-friendly representation of information.

3/data warhouse design:

In the development of the data warehouse, the structure of the output table was adjusted to align seamlessly with the specifications of a star schema.


branch_dim
bac_id varchar(255)
description varchar(255)?
students_dim
id int(11)
first_name varchar(512)?
last_name varchar(512)?
gender varchar(512)?
fact_table_student_performance
id int(11)
student_id int(11)?
spec_id varchar(3)?
bac_id varchar(512)?
level_id varchar(10)?
state_id varchar(5)?
date_id varchar(512)?
GPA double?
state_dim
state_id varchar(5)
description varchar(50)?
specialization_dim
spec_id varchar(3)
description varchar(50)?
date_dim
date_id varchar(512)
Day int(11)?
Month int(11)?
Year int(11)?
level_dim
level_id varchar(10)
description varchar(50)?

Fact Table:

1.Table Name: fact_table_student_performance

2.Attributes:

-id (Primary Key): Unique identifier for each record in the fact table.
-student_id: Foreign key referencing the primary key of the students_dim dimension.
-spec_id: Foreign key referencing the primary key of the specialization_dim dimension.
-bac_id: Foreign key referencing the primary key of the branch_dim dimension.
-level_id: Foreign key referencing the primary key of the level_dim dimension.
-state_id: Foreign key referencing the primary key of the state_dim dimension.

-date_id: Foreign key referencing the primary key of the date_dim dimension.

3.Measures:

-GPA: Measure representing the Grade Point Average.

Dimensions:

1.Branch Dimension:

Table Name: *branch_dim*

Attributes:

-bac_id (Primary Key): Unique identifier for each branch.

-description: Description of the branch.

2.Date Dimension:

Table Name: date_dim

Attributes:

-date_id (Primary Key): Unique identifier for each date.

-Day: Day component of the date.

-Month: Month component of the date.

-Year: Year component of the date.

3.Level Dimension:

Table Name: level_dim

Attributes:

-level_id(Primary Key): Unique identifier for each academic level.

-description: Description of the academic level.

4.Specialization Dimension:

Table Name: specialization_dim

Attributes:

-spec_id :(Primary Key): Unique identifier for each specialization.

-description: Description of the specialization.

5.State Dimension:

Table Name: state_dim

Attributes:

-state_id (Primary Key): Unique identifier for each state.

-description: Description of the state.

6.student dimension:

Table Name: students_dim

Attributes:

-id (Primary Key): Unique identifier for each student.

-first_name: First name of the student.

-last_name: Last name of the student.

-gender: Gender of the student.

4/Multidimensional Analysis:

- Tools: Microsoft SQL Server (SSAS) and Workbench Schema

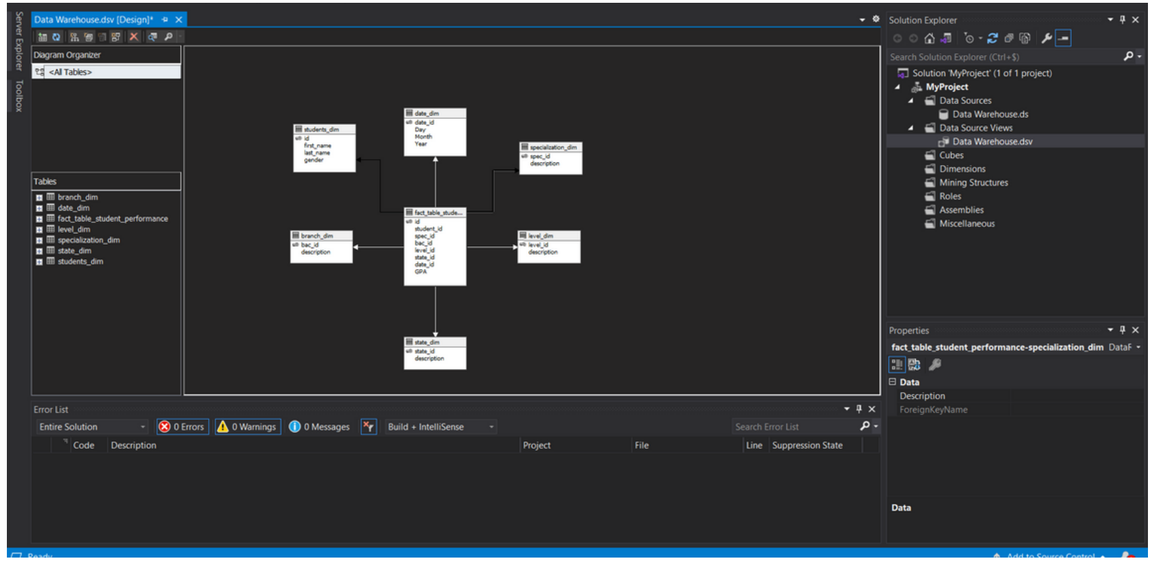
Initial Setup:

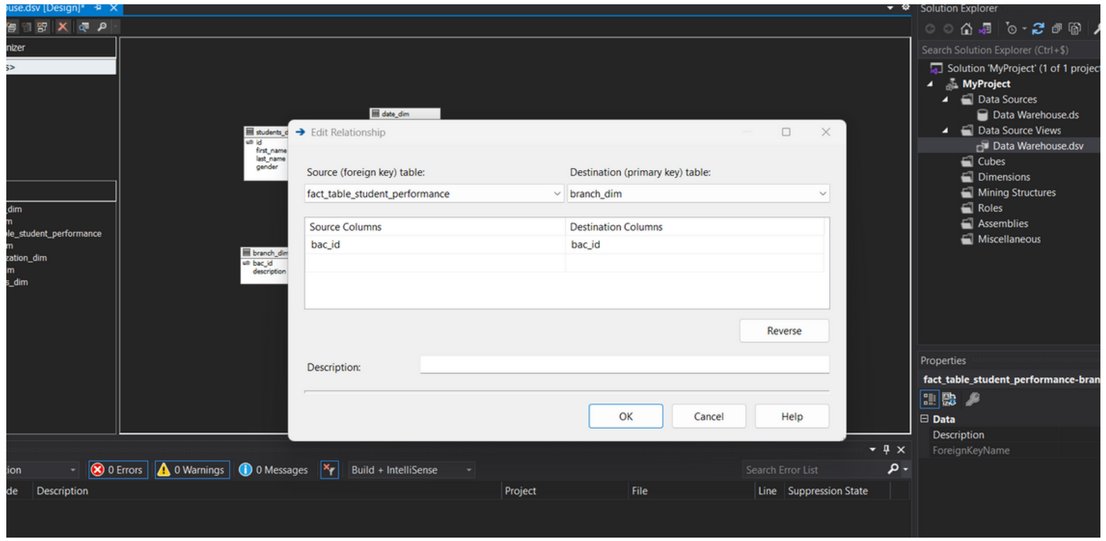
- Two Servers:
  - SSAS
  - SQL Server Management Software
- Integrated tables, defined dimensions, fact table, and measures.

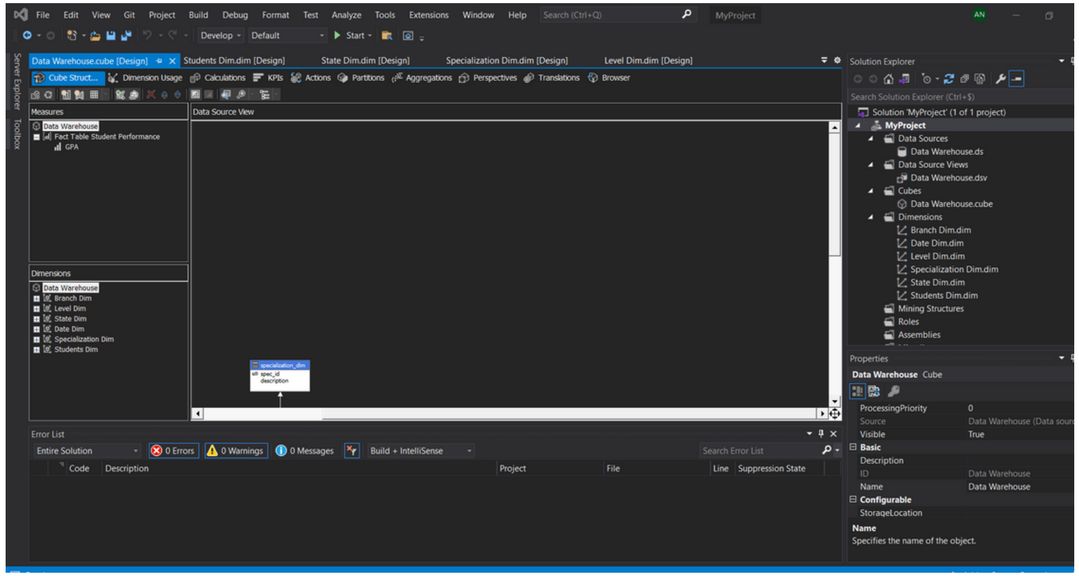
Screenshot:

Include a snapshot of the tables and dimensions setup.

4/Multidimensional Analysis:







4/Multidimensional Analysis:

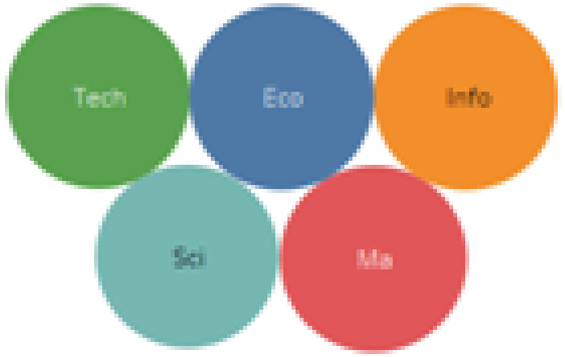
Transition and Implementation:

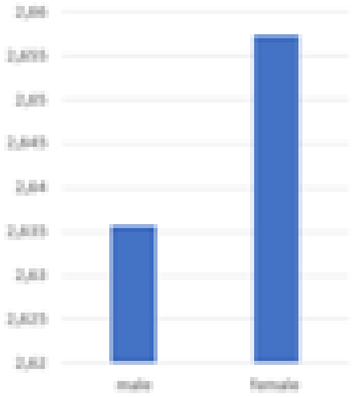
- Faced connection errors with SSAS.
- Transitioned to Workbench Schema.
- Created schema, fact table, and dimensions.
- Used MDX queries for analysis.

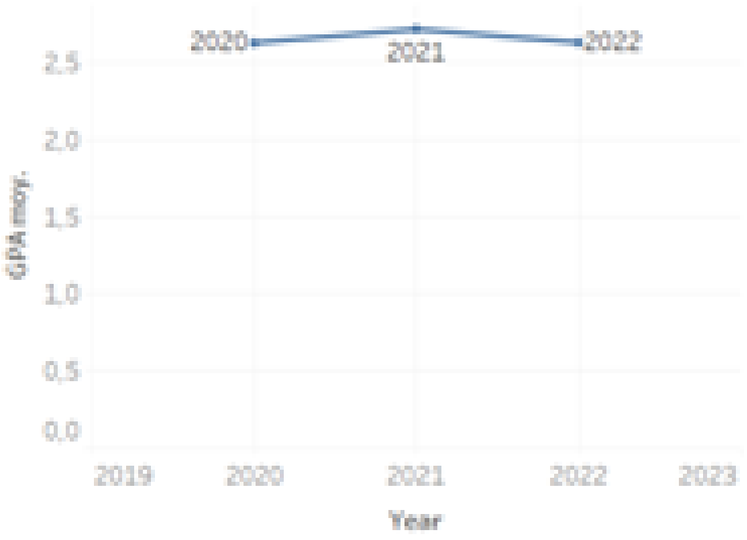
5/data visualisation:

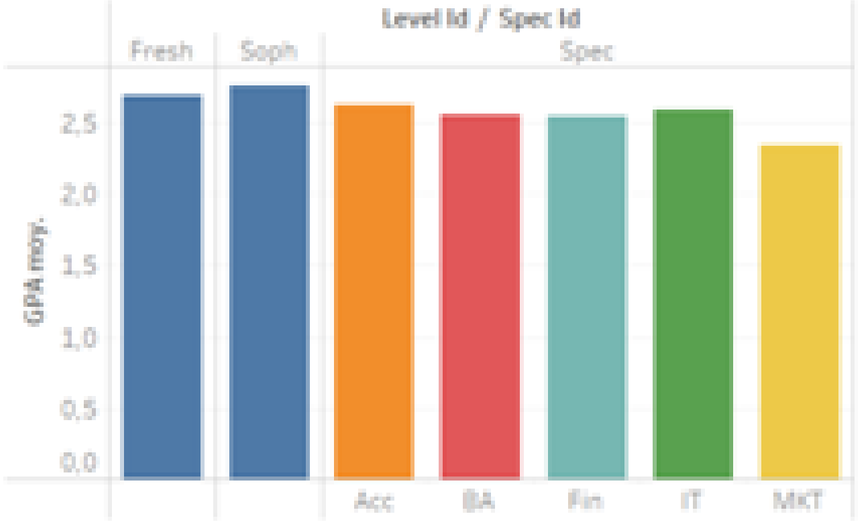
Following the establishment of the fact table and dimensions, SQL data was imported into Tableau for the purpose of constructing a dashboard. This dynamic dashboard serves as a comprehensive platform, presenting diverse insights garnered throughout the entire process, empowering users to visually explore and analyze the data with utmost effectiveness.

students performance dashboard

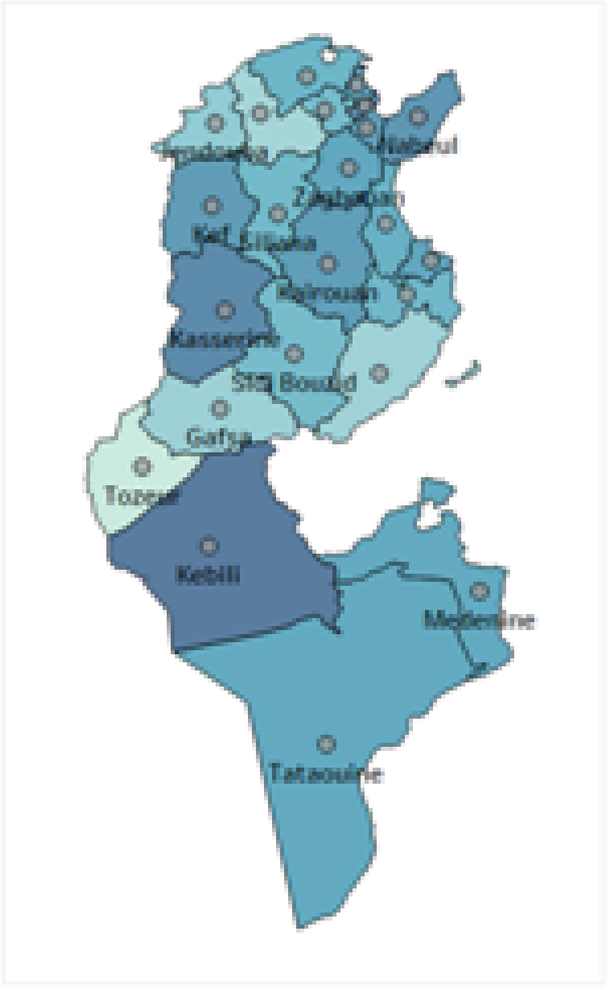








Average of GPA for each Spec Id broken down by Level Id. Color shows details about Spec Id.



Spec Id
- Acc
- BA
- Fin
- IT
- MKT

Bac Id
- Eco
- Info
- Ma
- Sci
- Tech

Avg. GPA
2.4326 2.8193

Conclusion:

Our business intelligence project journey was marked by triumphs and challenges, notably during the installation phase of software, servers, and interfaces. Critical issues arose, impacting timelines and strategies.

Challenges:

- Tool Selection: Starting with Talend for transformation, we encountered errors prompting a late transition to Python, affecting dataset adjustments.
- Connection Issues: Persistent problems with SQL server connections significantly hampered progress, especially during the critical OLAP phase.
- Mondrian Implementation: Installing Mondrian and applying procedures from the UVT website proved challenging, complicating the OLAP process.
- Documentation Gaps: Limited documentation for certain tools hindered efficient issue resolution within project timelines.

Streamlined Enhancements:

1.Practical Tool Training:
Enhancement: Conduct hands-on training for effective implementation of selected tools.
2.Project-focused Learning:
Enhancement: Introduce project-based learning for real-world skill enhancement.
3.Curated Learning Resources:
Enhancement: Provide tailored online resources for self-paced learning.
4.Extended Data Scope:
Enhancement: Expand data collection for more comprehensive insights.
5.Scalable Data Architecture:
Enhancement: Design scalable data architecture for evolving needs.

Reflection:

Despite challenges, the experience gained provides valuable insights. Addressing installation issues and refining tool selection and documentation approaches will contribute to more seamless business intelligence projects in the future. Our adaptability positions us well for future endeavors in decision-making systems.

Thank you!