

# **Requirements specification for recruitment business process**

## **1. General description of business process**

1. Once every half a year there is a recruitment process ongoing. Students use a dedicated website to choose their preferred courses of study and rank from those that they want to study the most and the least. After the deadline, each student has their grade calculated based on the A levels grades. For a specific course, students with the highest scores are selected until the course is full. For students that were unable to get on the list at their first preference course, score for next pick is calculated until the student is assigned to one of the courses or they ran out of preferred courses.

The increase in the number of students at the whole university at a level not less than 2,5% per semester compared to the previous one.

Median increase of number of students at each department 2,5% per semester compared to the previous one.

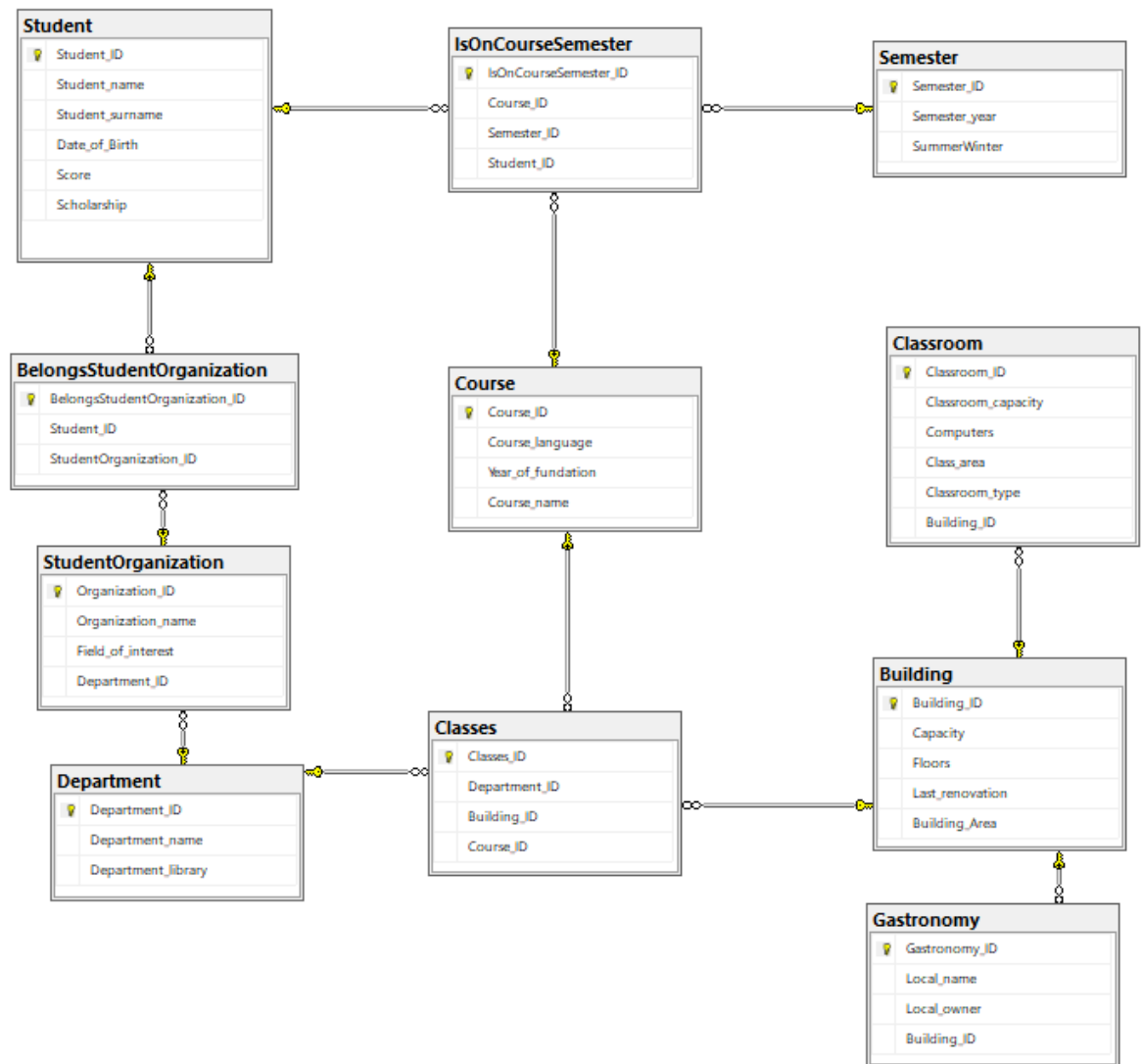
## **2. Typical questions:**

- What are the most popular courses?
- Which course has the fastest growth rate?
- Compare the number of students between departments.
- Compare the number of students at a given department with the number from last year.
- Compare the number of students at each course in a given department.
- What are the most popular departments?
- Which department lost the most students in comparison to the previous year?
- What is the least popular department?
- What is the least popular course?
- How many students are there on the least popular course?

### 3. Data:

All the data about the students are extracted from the system “MojaPG”. The system stores the information about the students and courses that they are enrolled in. Data about departments, capacity of their buildings and so on is stored in an Excel file.

### ERD



Every relationship above is one to many

ENTITY NAME	ATTRIBUTE	TYPE	DESCRIPTION
Student	Student and basic informations about him		
	Student_ID	Numercial	PK
	Student_name	String – 40 characters	First name
	Student_surname	String – 40 characters	Surname
	Date_of_birth	DateTime	Date of birth
	Student_address	String – 40 characters	Address
	Score	Numercial	Score from A levels
	IsOnCourseSemester_ID_FK	Numercial	Student is assigned to one course
	PESEL	11 digit number	Unique Pesel of a student
IsOnCourseSemester	Implementation of many-to-many relationship between Course and Semester. It means a given course conducted at a given time represented by Semester. It represents every semester of every year of a course.		
	IsOnCourseSemester_ID	Numercial	PK
	Course_capacity	Numercial	Number of possible spots on the course
	Candidates	Numercial	Number of people that applied for the course
	Course_ID_FK	Numercial	FK pointing to the entity Course
	Semester_ID_FK	Numercial	FK pointing to the semester in which the certain course was conducted.
Semester	Every semester that took place at the university. It means that every year there will be two semesters added to the database.		
	Semester_ID	Numercial	PK
	Semester_year	Numerical	Year when the semester was conducted
	SummerWinter	Bit	Information if the semester was winter or summer
Course (kierunek)	Basic information about the courses.		

	Course_ID	Numercial	PK
	Course_name	String – 40 characters	Name of course
	Course_language	String – 40 characters	In which language are classes conducted
	Year_of_fundation	DateTime	When was the first recruitment started.
BelongsStudentOrganization	Implementation of many-to-many relationship between Student and StudentOrganization. It means a given student belongs to a given organization.		
	BelongsStudentOrganization_ID	Numercial	PK
	Student_ID_FK	Numercial	FK pointing to the specific student
	StudentOrganization_ID_FK	Numercial	FK pointing to the entity StudentOrganization
StudentOrganization	Organization and the field of its interest		
	Organization_ID	Numercial	PK
	Organization_name	String – 40characters	Name
	Field_of_interest	String – 50 characters	Brief description of the main goals of the organization
	Department_ID_FK	Numercial	Organization is located at a department
Classes			
	Classes_ID	Numercial	PK
	Department_ID_FK	Numercial	Classes are taught on a given department
	Building_ID_FK	Numercial	Classes take place in a given building
	Course_ID_FK	Numercial	Classes are conducted for a given course
Department	Basic informations about Department concerning its name and whether it has library		
	Department_ID	Numercial	PK
	Department_name	String – 50characters	Name
	Department_library	Bit	Is there a library on a

			department
Gastronomy	Basic informations about Gastronomy concerning its name, type and owner		
	Gastronomy_ID	Numercial	PK
	Local_name	String – 40 characters	Name
	Local_type	String – 40 characters	Restaurant/buffet/cafe and so on
	Local_owner	String – 40 characters	Owner
	Building_ID_FK	Numercial	Local is in the building
Building	Basic informations about Building concerning its capacity, area, number of floors and the date of the last renovation		
	Building_ID	Numercial	PK
	Capacity	Numercial	Max number of students that can have classes in the building at the same time
	Floors	Numercial	Number of floors in the building
	Last_renovation	DateTime	When was the last renovation conducted
	Building_Area	Numercial	Are of the building in square meters
Classroom	Basic informations about classroom concerning its capacity, area and whether it is equipped in computers or not		
	Classroom_ID	Numercial	PK
	Classroom_capacity	Numercial	How much students can fit in a classroom
	Computers	Bit	Does the classroom have computers
	Classroom_area	Numercial	Are of the classroom in square meters
	Building_ID_FK	Numercial	Classroom is located in a building

## EXCEL

- Column A - Student ID
- Column B - Nationality
- Column C - Erasmus student (1 if Yes, 0 if No)
- Column D - IZP student (1 if Yes, 0 if No)
- Column E - Scholarship

Analytical problems:

*sum of all students* - depending on the query, we take Courses which are connected to a Semester of a given year or all courses at the university. Then, for each of them, we count the number of students enrolled into that course. The final number is the sum of all of those smaller sums for each course that we are interested in.

How do foreign and erasmus students influence the total growth in the university?

- 1. Show what percentage of all students, in the year with the highest number of students, were erasmus and foreign ones.**

*sum of all students* - taken only from courses that are connected to the semester with highest number of student

*year with highest number of students* - max value from sum of all students (only 'smaller' sums)

*erasmus students* - from EXCEL file we take IDs of all students with column C (Erasmus student) equal to 1, then we sum only those IDs that in ERD are connected to course that is on semester with highest number of students

*foreign students* - from EXCEL file we take IDs of all students with column B (Nationality) different from Polish, then we sum only those IDs that in ERD are connected to course that is on semester with highest number of students

- 2. Show the growth of students who graduated from highschool and those who graduated from technical school between year 2021 and 2022. (Requires additional data and change in business process - holding information about graduation)**

*sum of all students at the given course* - sum of students that are connected to the course that we are analyzing

*IZP students* - while

*Erasmus students* - EXCEL, column Erasmus

- 3. Show the average percentage of foreign students between Gdańsk University of Technology and other technology universities in the year of 2022. (Requires additional data but does not change business process)**

*sum of all students* - courses connected to the semester with attribute Semester\_year equal to 2022

*foreign students* - in EXCEL, we take IDs of all foreign students and check in ERD if they are attending course that is connected to Semester → Semester\_year = 2022

*number of students of other university* - data must be collected from public available data of the compared university

*foreign students of other university* - data must be collected from public available data of the compared university

- 4. Show student growth at the courses conducted in english between the year 2021 and 2022 and the course average for the whole university during the same time.**

*sum of all students* - for year 2021 and 2022, then calculate the growth number

*sum of all students* - for year 2021 and 2022, attribute

Course\_language in the entity Course has to be equal to English, then calculate the growth number

- 5. Show what percentage of total recruited students are those from Erasmus and whether they have a scholarship. Do the same for non-Erasmus students.**

*sum of all students* - for every year that we are analyzing

*foreign students* - we take all IDs of students with nationality other than Polish in the EXCEL

*sum of all foreign student* - calculated as before but with IDs taken from EXCEL

- 6. Show what percentage of all students, in the year with the lowest number of students, were not from Europe.**

*sum of all students* - taken only from courses that are connected to the semester with highest number of student

*year with lowest number of students* - lowest value from sum of all students (only 'smaller' sums)

*students not from Europe* - from EXCEL file we take IDs of all students with column B nationality of student, then we sum only those IDs that in ERD are connected to course that is on semester with lowest number of students

- 7. Show the percentage of Erasmus students and whether they have a scholarship. Do the same for non-Erasmus students.**

Erasmus students - from Excel

How infrastructure impacts growth of the number of students at the university?

- 1. Show average student growth between departments in which over 60% of buildings were last renovated after 2010 with the ones that do not satisfy that condition. Growth is calculated from the year 2021 to 2022.**

*departments with over 60% renovated buildings* - for every department we take all of the buildings that are connected to it and calculate the number of those which satisfy the criterion and those which don't. Lastly, the number of renovated buildings compared to the total number of buildings has to be higher than 60%.



*other departments - those departments that do not satisfy the above condition*

*date of renovation* - ERD, entity Building, attribute Last\_renovation  
*sum of students at the department* - we take all of the courses that are connected to a given department and then sum all of the students at a given course as we did in the first problem

**2. Show a difference in growth of students between buildings with and without any Gastronomy between the year 2021 and 2022.**

*buildings with gastronomy* - Buildings that have gastronomy connected to them

*sum of students having classes at a given building 2021 and 2022* - we take all courses that have some classes at a given building and sum the students enrolled to that course as in the previous problem. We do it for courses during the year 2021 and 2022.

**3. Show the average congestion (capacity divided by area) on a given department and growth in the number of the students there. Growth calculated from 2021 to 2022.**

*sum of all all students having classes at a given building* - we take all of the courses that have classes inside the given building and sum the number of students in those courses as before

*congestion in a building* - sum of all students having classes at a given building divided by Capacity of that building

*average congestion on a given department* - we calculate the congestion for every building that is connected to a given department and from those numbers we calculate the average congestion

- 4. Show the list of all departments, whether they have a library or not, the sum of all restaurants in the buildings where a given department have classes and the sum of student organizations located at those departments. For every single department, show the growth in the number of students between the year 2021 and 2022.**

*libraries at a given department* - attribute library in the entity Department

*sum of all restaurants* - we take all buildings where a given department have classes and sum all of the gastronomy that is inside that building

*sum of all student organization* - sum of the organizations that are connected to a given department

*sum of students at the department* - we take all of the courses that are connected to a given department and then sum all of the students at a given course as we did in the first problem. For courses during 2021 and 2022 only.

- 5. Show the percentage of classrooms with computers in the buildings of a given department and their growth in the number of recruited students at those departments between the year 2021 and 2022?**

*percentage of classrooms with computers for a department* - we take all of the buildings where a given department have classes and then we look at the attribute Computers to determine the percentage of classrooms with and without them

*sum of students at the department* - we take all of the courses that are connected to a given department and then sum all of the students at a given course as we did in the first problem. For courses during 2021 and 2022 only.

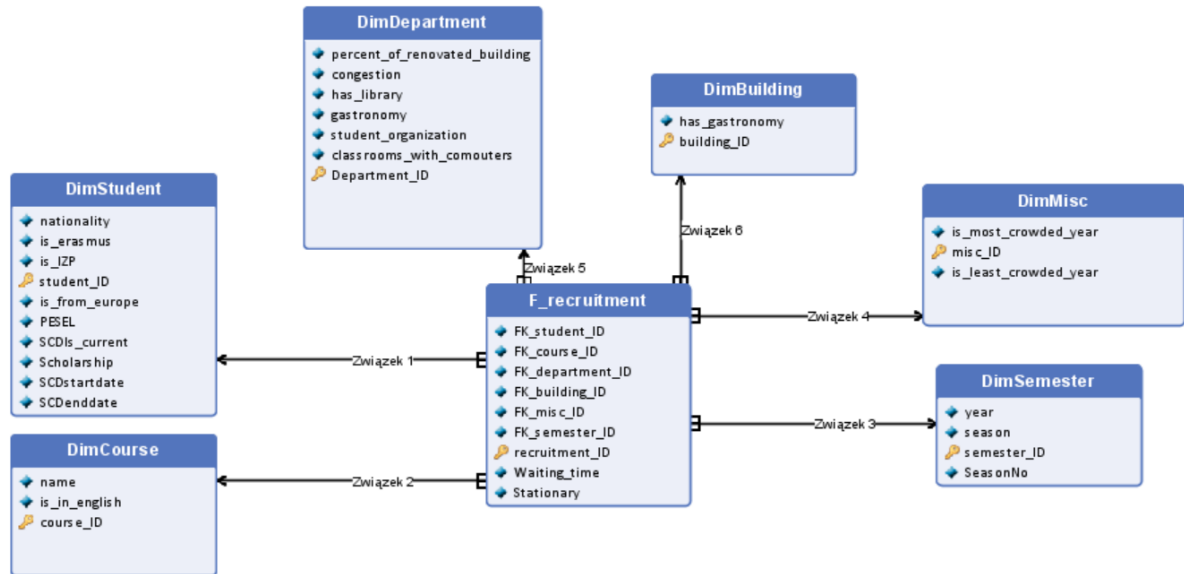


TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
F_recruitment(Fact table)	One tuple describes one student recruited at the university.		
	FK_student_ID	6 digits	FK DimStudents Student ID
	FK_course_ID	Numeric	FK DimCourse Course ID
	FK_department_ID	Numeric	FK DimDepartment Department ID
	FK_building_ID	Numeric	FK DimBuilding Building ID
	FK_misc_ID	Numeric	FK DimMisc misc ID
	semester_ID	Numeric	FK DimSemester Semester ID
	recruitment_ID	10 digits	Primary Composite Key
	Waiting_time	INT	Number of days
	Stationary	Varchar(44)	Submission of the application was stationary if True, Submission of the application was online if false
DimCourse(Dimension table)	One tuple describes one course(kierunek) conducted at university.		
	course_ID	Numeric	PK

	name	Varchar(30)	Name of the course
	is_in_english	Varchar(39)	The course is conducted in english if true, The course is not conducted in english if false
DimStudent(Dimension table)	One tuple describes one student at university.		
	student_ID	6 digits	PK(surrogate key)
	nationality	Varchar(20)	Nationality of student
	is_erasmus	Varchar(11)	Erasmus if true, Not Erasmus if false
	is_IZP	Varchar(7)	IZP if true,Not IZP if false
	is_from_europe	Varchar(15)	From Europe if true, Not from Europe if false
	Pesel	11 digits	BK
	SCDIs_current	Varchar(11)	current if true, not current if false (SCD2 implementation)
	Scholarship	Varchar(15)	Scholarship if true, No scholarship if false
	SCDstartdate	Date	Date of creating the tuple
	SCDenddate	Date	Date of invalidation of the tuple
DimDepartment(Dimension table)	One tuple describes one department at the university		
	Department_ID	Numeric	PK
	percent_of_renovated_building	Varchar(62)	Percentage of buildings that a given department has classes on and were renovated after the year 2010  The department has MORE than 60% of buildings renovated after 2010 if true,  The department has LESS than 60% of buildings renovated after 2010 if false
	congestion	Varchar(16)	Capacity of the department divided by the area of all buildings and classes that belong to that department  Variable can have three

			values: Small congestion- cong. $\leq 2$ AVG congestion- 5 > cong. > 2 Big congestion- cong. $\geq 5$
	has_library	Varchar(10)	Library if true, No library if false
	gastronomy	Varchar(16)	Number of all gastronomy places that belongs to the department  Variable can have three values: No gastronomy- gastr. = 0 One gastronomy- gastr. = 1 Many gastronomy- gastr. > 1
	student_organization	Varchar(26)	Number of all student organizations that are located on a given department  Variable can have three values: No student organization - stu. org. = 0 One student organization - stu. org. = 1 Many student organizations - stu. org. > 1
	classrooms_with_computers	Varchar(15)	Number of all classrooms with computers that belong to a given department  Variable can have four values: No classrooms - classrooms = 0% Some classrooms- 25% $\geq$ class. > 0 Many classrooms- 75% $\geq$ class. > 25% Lots classrooms- classrooms $\geq$ 75%
DimBuilding(Dimension table)	One tuple describes one building at the university.		
	building_ID	Numeric	PK
	has_gastronomy	Varchar(13)	Gastronomy if true, No gastronomy if false
DimMisc(Dimension table)	Tuple holds the value of the most crowded year at the university.		
	misc_ID	Numeric	PK

	is_most_crowded_year	Varchar(51)	Year with the highest number of recruited students.
	is_least_crowded_year	Varchar(50)	Year with the lowest number of recruited students.
DimSemester(Dimension table)	One tuple describes one semester.		
	semester_ID	Numeric	PK
	year	4 digits	Year
	season	Varchar(6)	Possible values: Winter Summer
	seasonNo	Int	1 if Winter, 2 if Summer

## Dimensional model

### Fact definitions

Fact 1 Recruitment: Recruitment of a specified student, recruited on a specified semester, on a specified course.

Fact Table:F\_recruitment

### Granularity:

- specified recruitment
- specified student, with specified nationality, erasmus and IZP status,
- specified department, with specified classes that can have computers or not, with specified gastronomy
- specified building
- specified course
- specified semester

### Measures and aggregation functions:

Number of recruitment facts - DISTINCT COUNT(1)

Number of Students - DISTINCT COUNT(Students)

Number of Erasmus Students - DISTINCT COUNT(is\_erasmus)

Number of foreign Students - DISTINCT COUNT(nationality)

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
Recruitment number	F_recruitment.recruitment_ID	Degenerate dimension
DimCourse	DimCourse	Dimension
Course name	DimCourse.name	Dimension attribute
Is course in english	DimCourse.is_in_english	Dimension attribute
DimStudent	DimStudent	Dimension
Student nationality	DimStudent.nationality	Dimension attribute
Is student erasmus	DimStudent.is_erasmus	Dimension attribute
Is student IZP	DimStudent.is_IZP	Dimension attribute
Is student from europe	DimStudent.is_from_europe	Dimension attribute
Unique PESEL of a student	DimStudent.PESEL	Dimension attribute
Tells if the record of a student is current one	DimStudent.SCDIs_current	Dimension attribute
Tells if the student has scholarship	DimStudent.Scholarship	Dimension attribute
Tells the date of creation	DimStudent.SCDstartdate	Dimension attribute
Tells the date of validity	DimStudent.SCDenddate	Dimension attribute
DimDepartment	DimDepartment	Dimension
Percentage of renovated buildings in the department	DimDepartment.percent_of_renovated_building	Dimension attribute
Congestion of the department	DimDepartment.congestion	Dimension attribute
Does department has a library	DimDepartment.has_library	Dimension attribute
Number of gastronomy in the department	DimDepartment.gastronomy	Dimension attribute
Number of student organizations in the department	DimDepartment.student_organizations	Dimension attribute
Number of classrooms with the computers	DimDepartment.num_of_classrooms_with_computers	Dimension attribute
DimBuilding	DimBuilding	Dimension
Does building have gastronomy	DimBuilding.has_gastronomy	Dimension attribute
DimMisc	DimMisc	Dimension
Most crowded year at the	DimMisc.is_most_crowded_year	Dimension attribute

university		
Least crowded year at the university	DimMisc.is_least_crowded_year	Dimension attribute
DimSemester	DimSemester	Dimension
Year of the semester	DimSemester.year	Dimension attribute
Season of the semester	DimSemester.season	Dimension attribute
Season number	DimSemester.SeasonNo	Dimension attribute
Recruitment Date hierarchy	1. DimSemester.year 2. DimSemester.season	Hierarchical dimension

Checking the feasibility of queries based on the multidimensional model:

- 1. Show what percentage of all students, in the year with the highest number of students, were erasmus and foreign ones.**

Measure: number of all students

Dimension: Misc(attribute: is\_most\_crowded\_year)

Dimension: DimStudent(attribute: is\_erasmus)

Dimension: DimStudent(attribute: nationality)

- 2. Show student growth at the courses conducted in english between the year 2021 and 2022 and the course average for the whole university during the same time.**

Measure: number of all students

Dimension: Misc(attribute: growth\_at\_english\_courses)

Dimension: DimCourse(attribute: is\_in\_english)

- 3. Show the growth of the number of foreign students over the last five years, show the growth in the number of total students during the same period. Growth is calculated based on the previous year.**

Measure: number of all foreign student

Measure: number of all students



- 4. Show what percentage of all students, in the year with the lowest number of students, were not from Europe.**

Measure: number of all students

Dimension: Misc(attribute: is\_least\_crowded\_year)

Dimension: DimStudent(attribute: is\_from\_europe)

- 5. Show what percentage of total recruited students are those from Erasmus and whether they have a scholarship. Do the same for non-Erasmus students.**

Measure: number of all erasmus students

Measure: number of all student

Dimension: DimStudent(attribute: scholarship)

- 6. Show average student growth between departments in which over 60% of buildings were last renovated after 2010 with the ones that do not satisfy that condition. Growth is calculated from the year 2021 to 2022.**

Measure: number of all student

Dimension: DimDepartment(attribute: percent\_of\_renovated\_building)

- 7. Show a difference in growth of students between buildings with and without any Gastronomy between the year 2021 and 2022.**

Measure: number of all student

Dimension: DimMisc(attribute: growth\_at\_buildings\_with\_gastro)

- 8. Show the average congestion (capacity divided by area) on a given department and growth in the number of the students there. Growth calculated from 2021 to 2022.**

Measure: number of all student

Dimension: DimDepartment(attribute: congestion)

- 9. Show the list of all departments, whether they have a library or not, the sum of all restaurants in the buildings where a given department have classes and the sum of student organizations located at those departments. For every single**

**department, show the growth in the number of students between the year 2021 and 2022.**

Measure: number of all student

Dimension: DimDepartment(attribute: has\_library)

Dimension: DimDepartment(attribute: num\_of\_gastronomy)

Dimension: DimDepartment(attribute:  
num\_of\_student\_organization)

**10. Show the percentage of classrooms with computers in the buildings of a given department and their growth in the number of recruited students at those departments between the year 2021 and 2022?**

Measure: number of all student

Dimension: DimDepartment(attribute: number\_of\_classrooms)

Dimension: DimDepartment(attribute:  
number\_of\_classrooms\_with\_computers)

TABLE NAME	COLUMN	SOURCE
F_recruitment	One tuple describes one student recruited at the university.	
	recruitment_ID	Composite key generated by database
	FK_student_ID	Foreign key to dimension table DimStudent. Based on student ID from ERD
	FK_course_ID	Foreign key to dimensional table DimCourse. Based on course ID from ERD
	FK_department_ID	Foreign key to dimensional table DimDepartment. Based on department ID from ERD
	FK_building_ID	Foreign key to dimensional table DimBuilding. Based on building ID from ERD
	FK_misc_ID	Foreign key to dimensional table DimMisc. DimMisc contains useful values calculated based on various ERD values.
	FK_semester_ID	Foreign key to dimensional table DimSemester Based on semester ID from ERD
	Waiting_time	Int value, represents the number of days that the student waited for their application processing.
	Stationary	Bit value, 0 if the application was submitted

		online, 1 if stationary
DimCourse	One tuple describes one course(kierunek) conducted at university.	
	course_ID	Primary key generated by database
	name	Name of the course taken from Course table in ERD
	is_in_english	TRUE if the course is conducted in English, FALSE if not. Value taken from table Course in ERD.
DimStudent	One tuple describes one student at university.	
	student_ID	Surogate key generated by database
	nationality	Nationality of a student taken from Excel file from column B
	is_erasmus	1 if the student is Erasmus, 0 if not, taken from Excel file from column C
	is_IKP	1 if student is on IKP, 0 if not, taken from Excel file from column D
	is_from_europe	1 if student comes from Europe, 0 if not, based on aggregation of nationality taken from Excel file from column B
	PESEL	Business key taken from PESEL attribute from the Student table in ERD
	SCDIs_current	"1" if information is current, otherwise "0" (SCD2 implementation).
	SCDstartdate	SCD implementation
	SCDenddate	SCD implementation
DimDepartment	One tuple describes one department at the university.	
	Department_ID	Primary key generated by database
	percent_of_renovated_building	Percentage of buildings that a given department has classes on and were renovated after the year 2010. Value calculated based on attribute last renovation in table Building in ERD.
	congestion	Capacity of the department divided by the area of all buildings and classes that belong to that department. Calculated based on Area and number of students at a given department taken from tables Building and Course.
	has_library	TRUE if a given department has a library, FALSE if not. Value taken from table Building from ERD.
	gastronomy	Number of all gastronomy places that belong to the department. Value calculate based on the

		attribute gastronomy in table building in ERD
	student_organization	Number of all student organizations that are located on a given department
	num_of_classrooms_with_computers	Number of all classrooms with computers that belong to a given department
DimBuilding	One tuple describes one building at the university.	
	building_ID	Primary key generated by database
	has_gastronomy	TRUE if a given building has a gastronomy, FALSE if not. Value taken from table Building from ERD.
DimMisc	Tuple holds the value of the most crowded year at the university.	
	misc_ID	Primary key generated by database
	is_most_crowded_year	Year with the highest number of recruited students. Value calculated based on the amount of students in ERD.
	is_least_crowded_year	Year with the lowest number of recruited students. Value calculated based on the amount of students in ERD.
DimSemester	One tuple describes one semester.	
	semester_ID	Primary key generated by database
	year	Year in which the semester was conducted, taken from attribute year from Semester table in ERD
	SeasonNo	Number of a given season
	season	Season in which the semester was conducted, taken from attribute season from Semester table in ERD. Possible values: Winter Summer