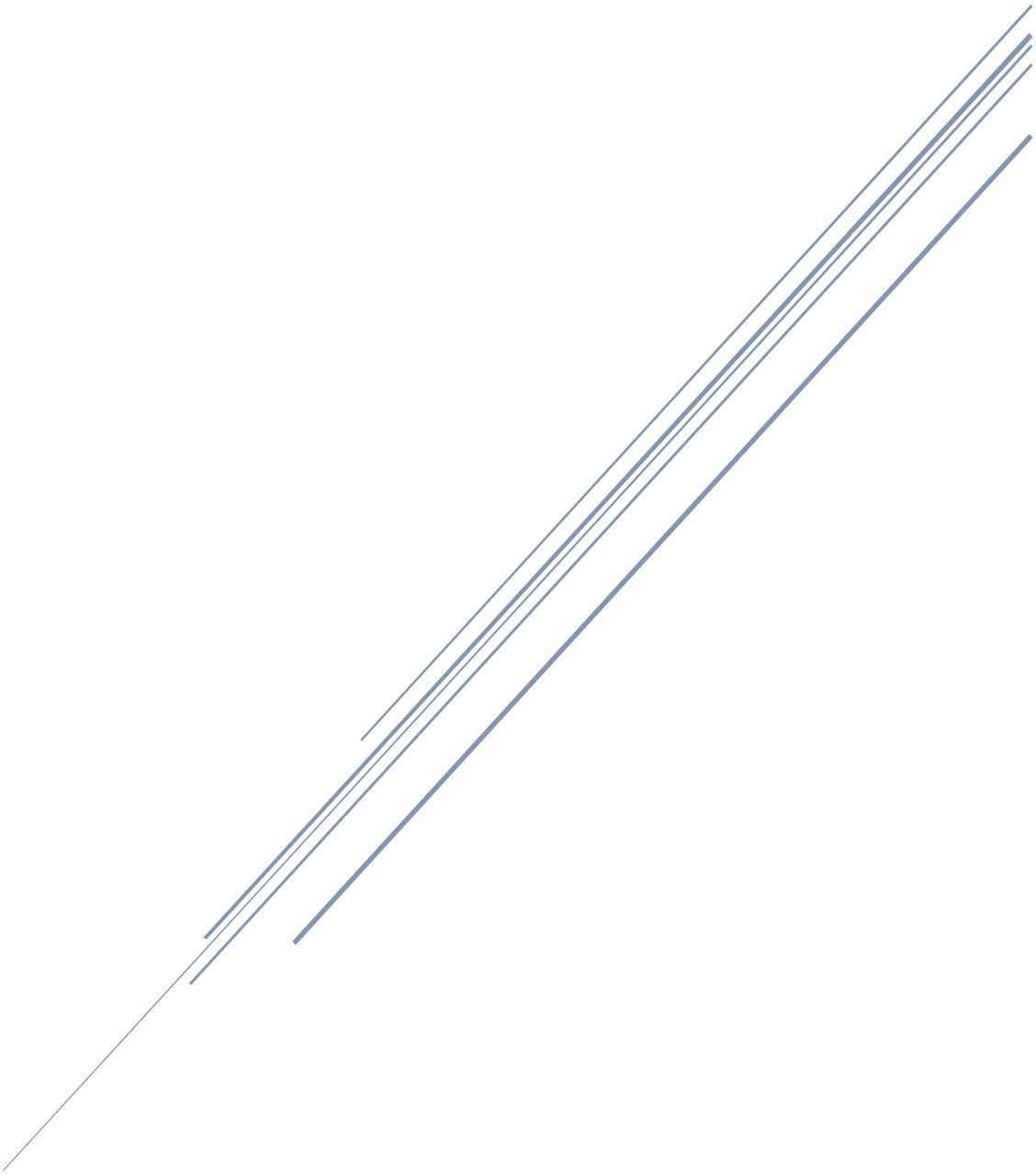
Data Visual

PLANNING AND CREATING STATIC VISUALISATIONS

Word count: 2300-2400 approx. Exclusive of tables, figures, and reference list



Jayden Dzierbicki
MA5830

Contents

Summary

Target Audience

The target audience for this assignment will be parents of students who are reviewing a final report of the students' performance across the course – we will assume the parents and school are in ACT, Canberra.

Questions to be addressed

1. The opening page of the report will provide a summary of grades achieved in line with the Australian Grading System competencies ("Grading System in Australia, Australian grading system", 2021)
2. The body of the report will include details of each individual subject, for the purpose of this report we will focus on one subject and the same logic can be applied.
3. The end of the report will demonstrate how the individual student performed relative to other students overall in the curriculum.

Audience Assumptions

- **General Assumptions:** We will have a range of parents from both white collar, blue collar and possibly unemployed with low levels of education. ABS data suggests that Canberra has a high relative median weekly household income compared to other states in Australia ("2016 Census QuickStats: Canberra", 2016), suggesting high levels of education or professional roles. We will assume that parents are educated at a year 12 level at a minimum ("2016 Census QuickStats: Canberra", 2016).
- **Data Assumptions:** Based on the general assumptions made we will assume that parents at a minimum have had exposure to high school level to statistics, with a majority of Canberrans working in the Australian Public Service we will also assume that they have exposure to data. It is assumed that parents will understand **median values, mean values, maximum values and minimum values,** though concepts such as **confidence intervals** and **standard deviations** may be abstract to some.
- **Visualisation Assumptions:** It will be assumed that parents are well verse with commonly used visuals such as bar charts, pie charts, scatterplots and visuals; and may not be literate in complex statistical visualisations such as boxplots.
- **Interaction and Time Assumption:** It will be assumed that parents are time rich in reviewing the students final report and will have the ability to research how to read an abstract visualisation if deemed necessary.

Data Abstraction

The dataset supplied was located in the assessment 2 folder and comprised of three variables summarising eight students' grade (assumed to be final) across nine individual subjects, the subjects suggest either a collage level (year 11 and 12) with electives or university level, we will assume to be a collage level, though this highlights a limitation in the data through lack of information supplied as the analysis would vary depending on the above.

The data was supplied in PDF format, to ensure useability in programs such as Tableau the data was copied into an appropriate format in excel, this resulted in three columns with multiple rows. The columns created include Student, grade and subject, the file was saved locally and used in Tableau. Table 1 summarises the data used, along with new variables created in Tableau, not all summarises produced are used in the visualisations as this was done in the exploratory stage.

Table 1: Summary of variables used in assessment
**Equation located in appendix 1 table 5*

Variable name	**Data Type**	**Summary**			**Visual question**
Student {original}	Categorical: Nominal	Levels = 8	**	**	1,2,3
Subject {original}	Categorical: Nominal	Levels = 9	**	-	1,2,3
Grade {original}	Numeric: Ratio	Mean = 61	Range = 23 - 96	Mode = 45	1,2,3
Competency* {Calculated}	Categorical: Ordinal	Levels = 6	Mode = Satisfactory (24)	**	1
LOD_Avg_grade* {Calculated}	Numeric: Ratio	Mean = 61.67	**	**	3
Grade Physics {Calculated}	Numeric: Ratio	Mean = 67.5	Range = 34 - 90	Median = 69	2

Question 1

Visualisations Intent (why)

The first page of the report is to provide a summary of grades achieved for a student in a course, the following intent and response was proposed.

- **Intent 1**: Provide a summary of the final grade a student achieved in a class aligned with the Australian Grading System, excellent, satisfactory ect – a summary in relation to ordinal categorical values.
- **Intent 2:** Only demonstrate performance relative to an individual student.
- **Intent 3:** Summaries the final raw grade achieved in each course.
- **Intent 4:** Demonstrate the grades in a cascade effect, that is highest results appear first.
- **Response:** Provide a quick summary for parents to set the scene and expectation of more detailed visualisations.

Data description (what)

The data attributes shown in table 2 are required for the following visual, equations on how the attributes were derived can be located in appendix A table 5 with further summary statistics located in Table 1. Note that the data lacks other attributes such as attendance and student engagement which could explain grades below average.

Table 2 – Data attributes for Question 1

Attribute	**Type**	**Intent**	**Channels**	**Gestalt Principle**
Subject {original}	Categorical: Nominal	Intent 1		
Grade {original}	Numeric: Ratio	Intent 3	Magnitude channel – colour saturation	Similarity
Competence {derived}	Categorical: Ordinal	Intent 1		Enclosure

Visualisation selection and aesthetics (how)

Tables allow for interaction with ones verbal system as the recipient will tend to read them and they are deemed appropriate for a mixed audience (Knaflic, 2015). Based on table 2 it is evident that we are wishing communicating mixed data types, such as numeric ratio and categorical ordinal data; tables allow for effective communication of different units of measure (Knaflic, 2015).

- **Similarity:** Objects of similar colour are perceived to belong to the same group and can be leveraged in tables to help draw an audience's eyes in the direction we want them to focus (Knaflic, 2015), this suggest we are able to organise our numeric ratio for grades in either descending or ascending order – the use of saturation allows us to achieve this.

- **Enclosure:** When objects are physically enclosed together this suggests they belong to a group (Knaflic, 2015), by grouping subjects into their competencies we can exploit this property as seen in the left side of figure 1.

Aesthetics

- **Remove clutter:** Applying the Gestalt Principle we are able to control for clutter through ensuring that all elements on the graph or table add value and do not contribute to increase cognitive load. Through enclosure we are also able to reduce cognitive load by removing excessive text as seen in figure 1 and excessive borders which contribute to cognitive load (Knaflic, 2015).
- **Eye Beat Memory:** This rule of thumb states that if information can be presented simultaneously it has a lower cognitive load then consulting our working memory (Munzner & Maguire, 2015), figure 1 allows us to present information simultaneously, reducing cognitive load.

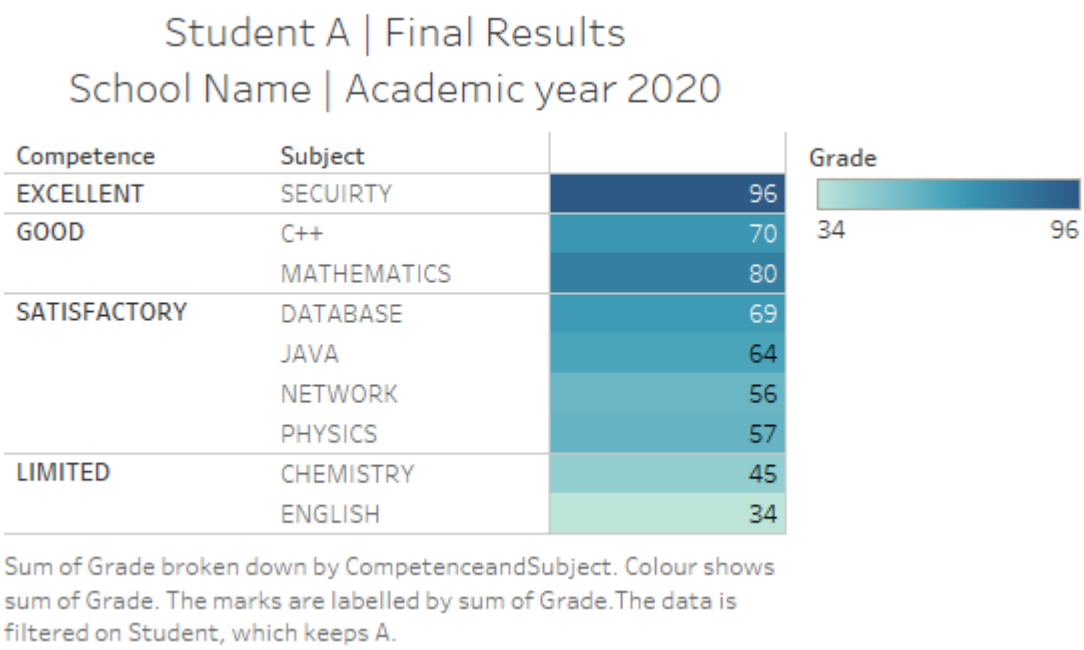


Figure 1: Static visual for question 1

Review:

The visualisation presented in figure 1 summarise a student's final grade and competency achieved in their course for the entire academic year, this summary is able to cater for a wide audience through the use of tables and achieves all the desired intents.

A graph of sorts was considered but during the exploratory stages was considered too noisy and had high levels of cognitive load for an otherwise simple solution as presented in figure 1.

The visual presented in figure 1 also adheres to elicit the desired response in a relative fast way, it allows the audience to quickly see the number of competencies achieved, as well as the final grade in a non-complex way.

The visual presented may cause some confusion to an extent, as the third column lacks a descriptive header and instead relays on the key to highlight what that column is, this might contribute slightly to increase cognitive load; future reports may benefit from minor adjustments as such noted and through the exclusion of an automatically generated descriptive text to more layman's terminology.

Question 2

Visualisations Intent (why)

The body of the report will include detailed summarise of a student's performance in a class, for the purpose of this assessment we will create one summary of a student A's performance in mathematics relative to other students – note the same logic will apply to other subjects and students.

- **Intent 1**: Convey a student's performance relative to other students in a class, the concept of percentiles is linked to quantiles.
- **Intent 2:** Ensure other student's information is non-identifiable.
- **Intent 3:** Demonstrate the minimum, maximum and median grade in a class.
- **Response:** The initial response may be some slight confusion; this is because we want to convey lots of information in a succinct manner. This is allowed because we have assumed parents have ample amount of time when reviewing a student's report and information will be provided to students on how to interpret the visual.

Data description (what)

The data attributes used for the data visual are shown in table 3, whilst the summary of statistics can be located in table 1. It should be noted that whilst no indication of sample size is given for a box plot it is best to have a sample size of at least 20 (Ngo, 2018), highlighting a possible limitation in our data sample size (n=8).

Table 3 – Data attributes for Question 2

Attribute	Type	Intent	Channels	Gestalt Principle
Subject {original}	Categorical: Nominal	Intent		
Grade {raw}	Numeric: Ratio	Intent	Magnitude channel – Position on common scale	Connection
Student {original}	Categorical: Nominal	Intent	Identity channel – colour hue	Similarity

Visualisation selection and aesthetics (how)

The type of visual elected based on the intent is a boxplot as it highlights multiple values associated with a 5 number summary which is attributed with the intent of the report. Given that the target audience is time rich it was deemed appropriate despite the target audience being unfamiliar with this data visual. When electing to use non-common visuals that the target audience may not be familiar with it is important to take extra care in making them both accessible and understandable (Knaflic, 2015), this can be achieved through the inclusion of a 'how to read a boxplot' summary at the start of the report.

- **Similarity:** Objects of similar colour are perceived to belong to the same group and can be leveraged in data visuals to help draw an audiences eyes in the direction we want them to focus (Knaflic, 2015), with the intent of identifying a specific student we are able to utilise colour to achieve this.

Aesthetics (how)

- **Remove clutter:** Applying the Gestalt Principle we are able to control for clutter through ensuring that all elements on the graph or table add value and do not contribute to increase cognitive load. Through similarity we are also able to reduce cognitive load by drawing the target audience's attention to only their child using neutral colours to reduce cognitive load (Knaflic, 2015). We also employ the 'get it right in black and white' rule of thumb which complements our similarity principle by encoding a luminance channel on student A whilst keeping the rest of the visual in black and white (Munzner & Maguire, 2015).
- **Overview First, Zoom and Filter, Details on Demand/Lack of visual order:** The intent of figure 2 is to provide a vis idiom that provides an overview of a student's performance overall in a particular course whilst showing all the data simultaneously without the need of consulting multiple sources. The inclusion of such information may assist parents in deciding if they wish to attend parent interviews or request more information about the data in line with this rule of thumb which states we can retain a large amount of information in some circumstances (Munzner & Maguire, 2015). This also allows us to effectively code our mark in line with the lack of visual order Gestalt Principle, further reducing cognitive load (Knaflic, 2015).

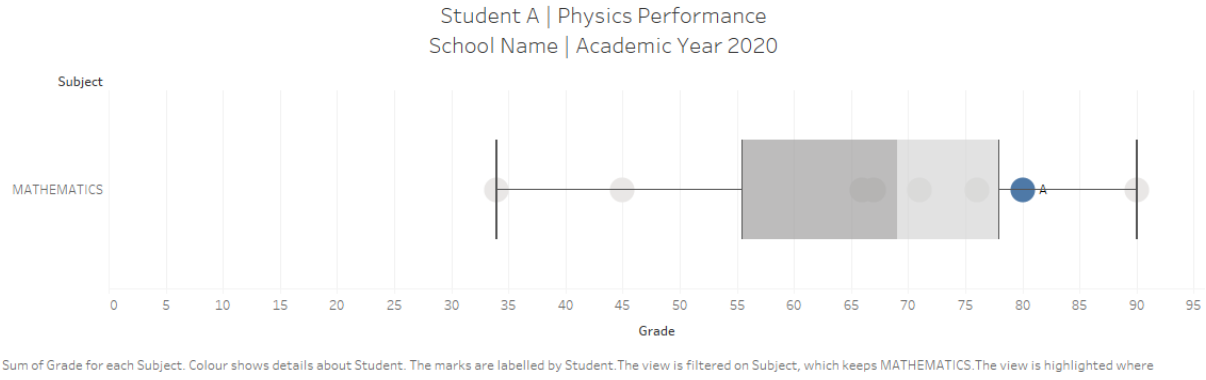


Figure 2: Static visual for question 2

Review:

Whilst the boxplot conveys information we deem the user to be literate in based on the data assumptions, a boxplot is not a commonly used visual and thus the audience may be initially confused by it. Though due to the time assumption we deemed it appropriate use of data visual if an instruction is provided at the start of report highlighting how to read a boxplot.

The boxplot otherwise achieves all the intents it set out to including the max grade, min grade, most frequent grade and student A's performance whilst also including extra information should the parent wish to delve deeper into the students' performance.

Another visual was created exploiting the analytics features of Tableau to create a data visual which included only the minimum, average and maximum grade achieved in the course for said student, whilst this might have been more appropriate for our target audience, we elected to still employ the boxplot due to the time rich assumption and to convey more information to our audience.

Question 3

Visualisations Intent (why)

The final part of the report will allow the parents to understand how the student performed on average against the cohort and recognising success.

- **Intent 1**: Demonstrate how an individual student performed in each subject relative to the average grade achieved by all students across all subjects.
- **Intent 2:** Highlight subjects in which the student did above average overall.
- **Response:** The desired response is on that quickly recognise the subjects in which student A did well in and was above the average grade.

Data description (what)

The data attributes used for the data visual are shown in table 4, whilst the summary of statistics can be located in table 1 and the LOD equation located in appendix A table 5. It should be noted based on table 1 and figure 2 that the grade data is skewed, this suggest the use of a median value might have been more appropriate in figure 3 (Sharma, 2019).

Table 4 – Data attributes for Question 3

Attribute	**Type**	**Intent**	**Channels**	**Gestalt Principle**
Subject {original}	Categorical: Nominal	Intent		
Grade {raw}	Numeric: Ratio	Intent	Grade, magnitude channel – position on common scale	Similarity
Student {original}	Categorical: Nominal	Intent		
LOD_Avg_Grade	Numeric: Ratio	Intent	Identity channel – Colour hue	Enclosure

Visualisation selection and aesthetics (how)

The benefits of bar charts is that they are common and less of a learning curve for a diverse audience (Knaflic, 2015) such as parents. For this visual we have elected for the horizontal bar chart as it is thought to be extremely easy for a user to read, especially when category names have long string text (Knaflic, 2015).

- **Similarity:** Objects of similar colour are perceived to belong to the same group and can be leveraged in data visuals to help draw an audience's eyes in the direction we want them to

focus (Knaflic, 2015), we exploit this principle by encoding grade data a certain colour to reduce cognitive load and draw

Aesthetics

- **Remove clutter:** Clear contract is a signal for our audience to help them focus and draw their attention on particular traits, it is thought that the lack of clear contract to be a form of visual clutter (Knaflic, 2015), this is achieved by the use of blue bars against grey bars.
- **Zero baseline:** The inclusion of a zero baseline is crucial for bar charts and can be a common mistake as the audience will interpret the information based on the length of the bar from a baseline (Knaflic, 2015), this can often skew and misguide the information otherwise.

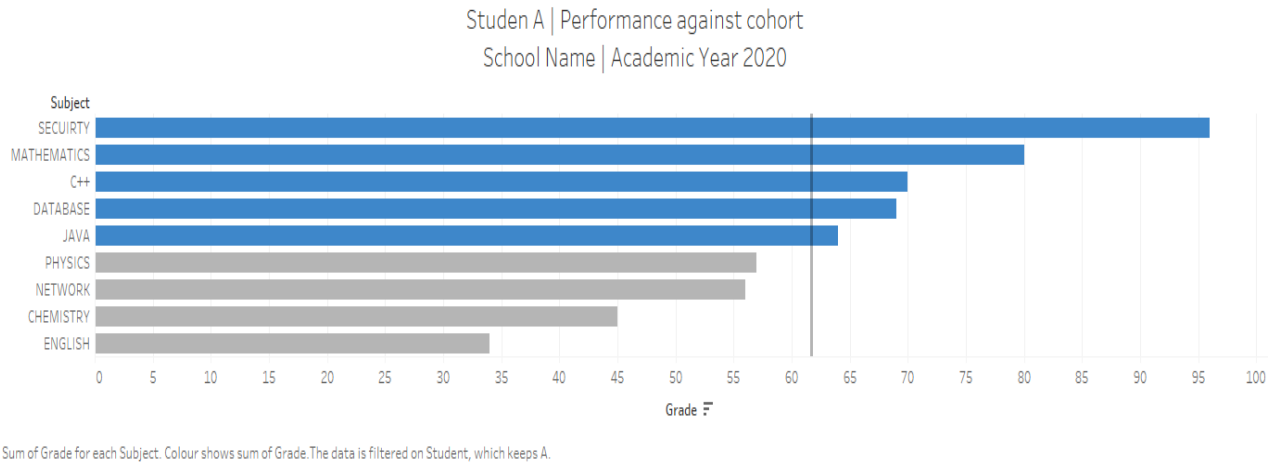


Figure 3: Static visual for question 3

Review:

The visualisation presented in figure 3 demonstrates the desired intent allowing for a broad audience to easily identify subjects in which a student performs above average in relative to the average grade achieved by the cohort.

A vertical bar chart was considered for this visual, though many of the categorical names were cut off due to their character length, resulting in increase cognitive load. It was also considered to superimpose the average grade achieved in each grade over the respective bar, but due to a small sample size of grades in each class and the increase visual load it was considered not appropriate in this situation based on the desired intent.

A future visual may benefit from the inclusion of how a student performed relative to the median grade as an alternative and the inclusion of values superimposed into the bar to reduce cognitive load as currently in figure 3 the parent is required to go down from the bar (y-axis) to find the corresponds grade value (x-axis).

Reference

Grading System in Australia, Australian grading system. (2021). Retrieved 16 September 2021, from https://www.australiaeducation.info/education-system/grading-system.html

Knaflic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professiona*. John Wiley & Sons.

Munzner, T., & Maguire, E. (2015). *Visualization analysis & design*.

Ngo, L. (2018). More on how to compare box plots - BioTuring's Blog. Retrieved 13 September 2021, from https://blog.bioturing.com/2018/05/22/more-on-how-to-compare-box-plots/

Sharma, R. (2019). Skewed Data: A problem to your statistical model. Retrieved 14 September 2021, from https://towardsdatascience.com/skewed-data-a-problem-to-your-statistical-model-9a6b5bb74e37

2016 Census QuickStats: Canberra. (2016). Retrieved 10 September 2021, from https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/CED801

Appendix

Table 5: Data transformations performed in Tableau

Variable	Tableau equation
Competency	IF [Grade] >= 85 THEN 'EXCELLENT' ELSEIF [Grade] >= 70 THEN 'GOOD' ELSEIF [Grade] >= 51 THEN 'SATISFACTORY' ELSEIF [Grade] >= 31 THEN 'LIMITED' ELSEIF [Grade] >= 26 THEN 'VERY LOW' ELSE 'FAILED' END
LOD_AVG_grade	{ FIXED : AVG([Grade]) }