

Can the Overall Satisfaction Score of a University Course be Modeled Against Student Survey Questions?

Motivation

The course lead for Food Science at “Super Worth It University” has approached me, a Data Science student at the university, to help them improve the poor student satisfaction score of their course. They want to know what aspects of the course they should focus on to improve the overall satisfaction score obtained from students.

Objective: Explore whether the overall satisfaction score of a course can be modeled against other survey questions/course statistics.

Goal: Produce a regression model that can best interpret a relationship for overall satisfaction and a model that can best predict overall satisfaction.

Methodology

Data source: <https://discoveruni.gov.uk/> → For both target and feature variables

- 1 Data scraping with Selenium and BeautifulSoup
- 2 Data storage with SQLite database and SQLAlchemy
- 3 EDA and regression modeling with Pandas and SKLearn
- 4 Results visualisation with Matplotlib and Seaborn

Methodology

47,342 Courses Total

Training and Validation Set (80%)

Test Set (20%)

Methodology

teaching_1

teaching_2

teaching_3

teaching_4

opportunities_1

opportunities_2

opportunities_3

feedback_1

feedback_2

feedback_3

feedback_4

support_1

support_2

support_3

management_1

management_2

management_3

resources_1

resources_2

resources_3

community_1

community_2

voice_1


voice_2

voice_3

voice_4

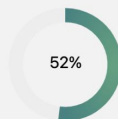
Methodology

Assessment and feedback

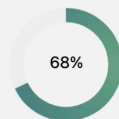
Source: [NSS \(National Student Survey\)](#) 

Data from 90 students
(72% of those who were asked)

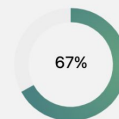
All final year students are asked to complete
the National Student Survey



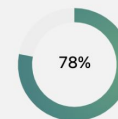
The criteria used
in marking have
been clear in
advance



Marking and
assessment has
been fair



Feedback on my
work has been
timely



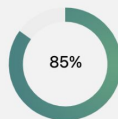
I have received
helpful comments
on my work

Academic Support

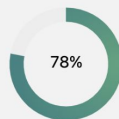
Source: [NSS \(National Student Survey\)](#) 

Data from 90 students
(72% of those who were asked)

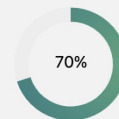
All final year students are asked to complete
the National Student Survey



I have been able
to contact staff
when I needed to



I have received
sufficient advice
and guidance in
relation to my
course

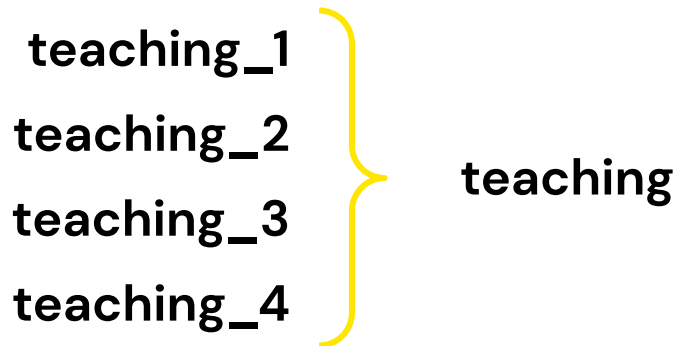


Good advice was
available when I
needed to make
study choices on
my course

From discoveruni.gov.uk

Methodology

Some function



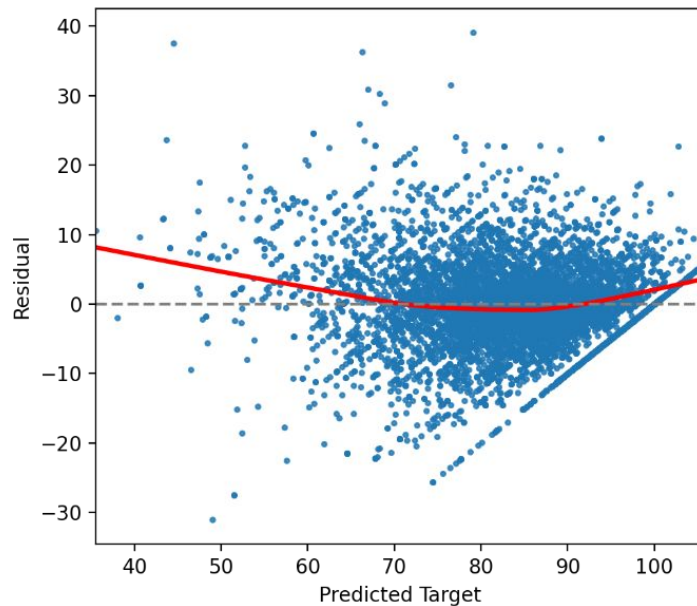
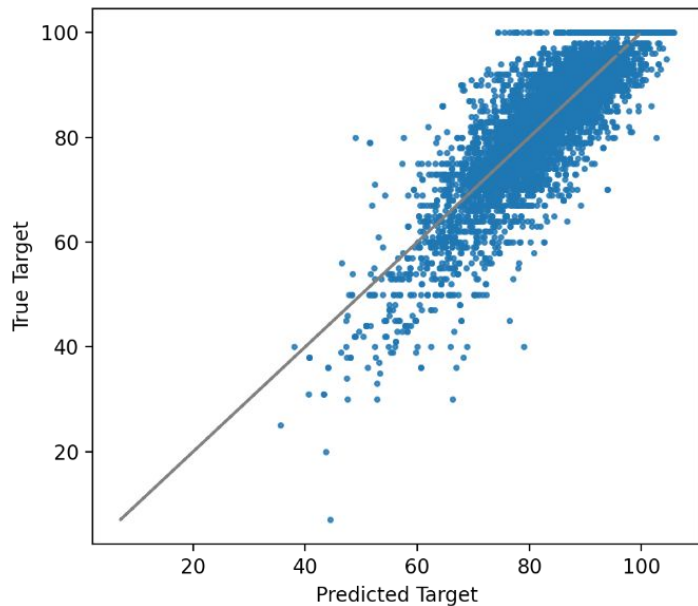
teaching_1
teaching_2
teaching_3
teaching_4

teaching

Functions tried: sum, mean, min, max

Results

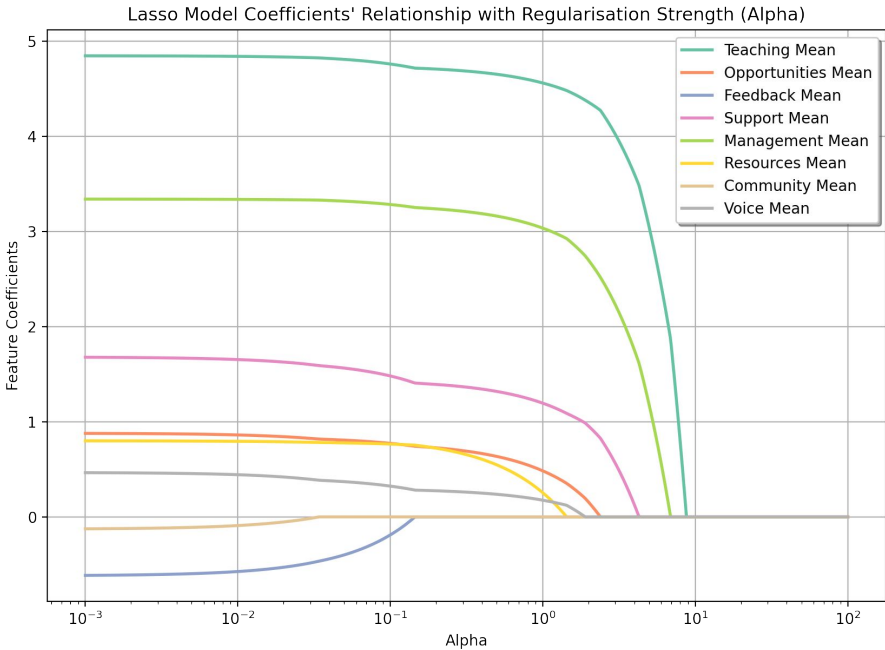
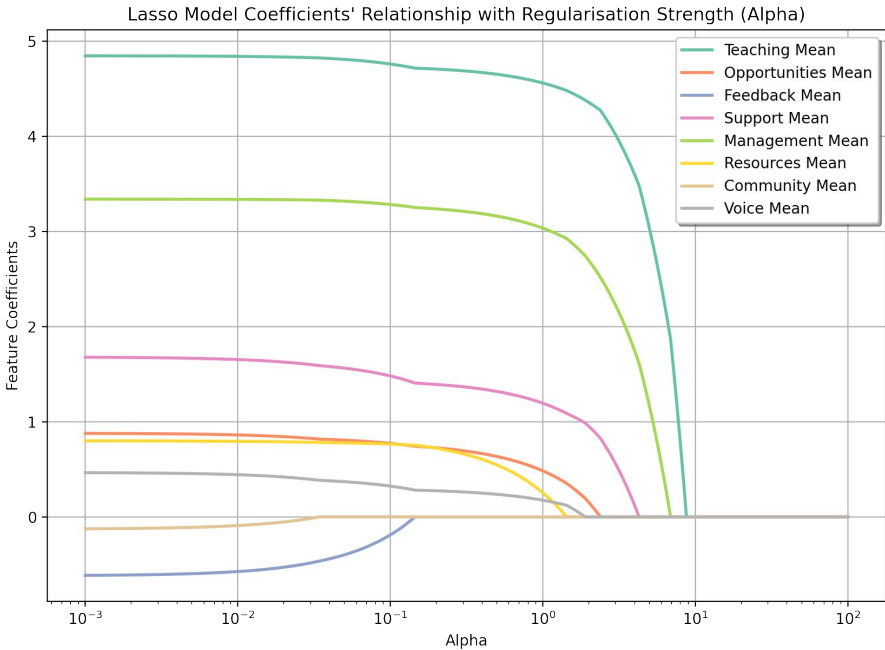
Linear Regression Model



Test Set Results: $R^2=0.7077$ MAE=4.53

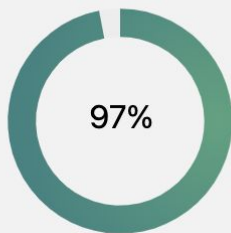
Results

Lasso Path

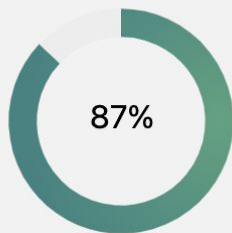


Results

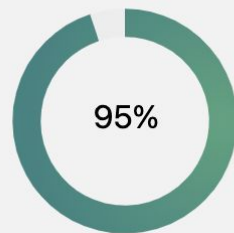
Teaching Questions



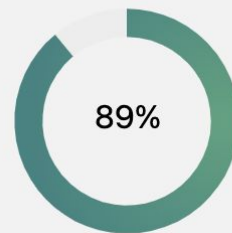
Staff are good at explaining things



Staff have made the subject interesting



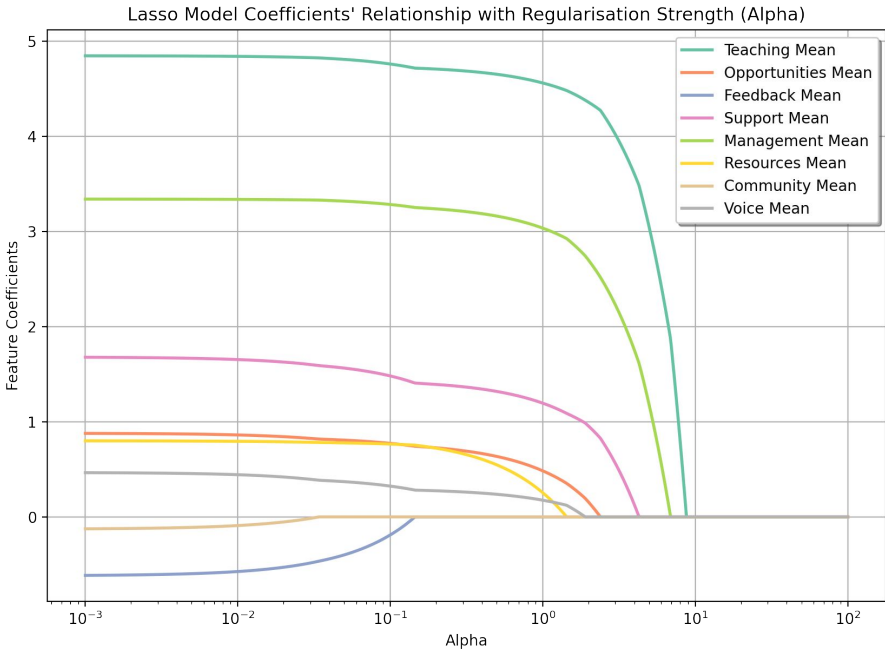
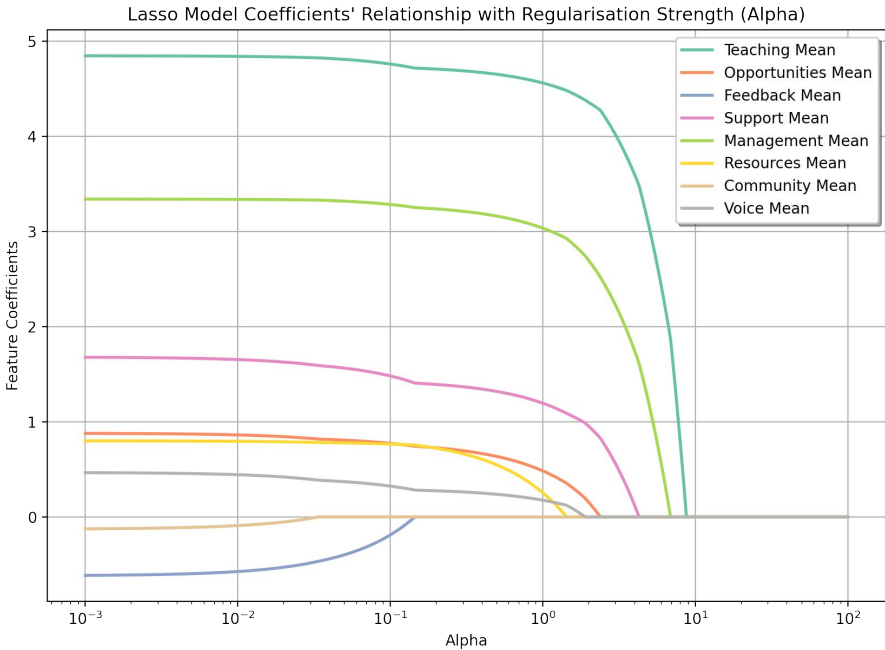
The course is intellectually stimulating



My course has challenged me to achieve my best work

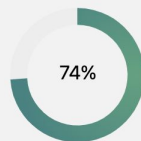
Results

Lasso Path

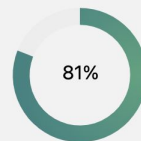


Results

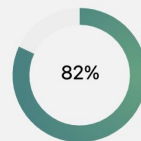
Management Questions



The course is well organised and running smoothly

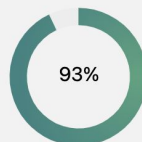


The timetable works efficiently for me

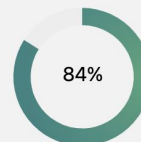


Any changes in the course or teaching have been communicated effectively

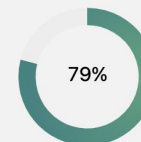
Support Questions



I have been able to contact staff when I needed to



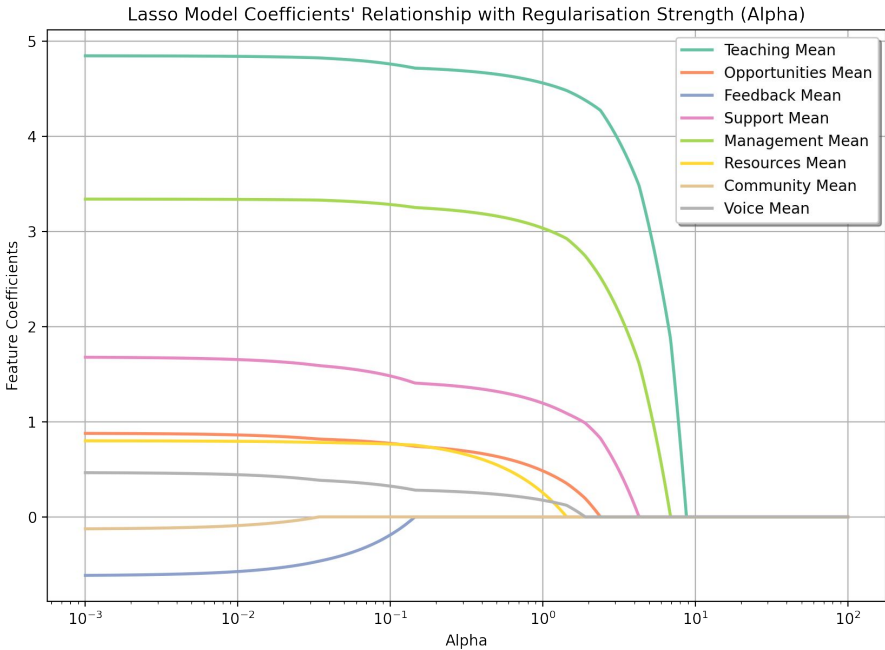
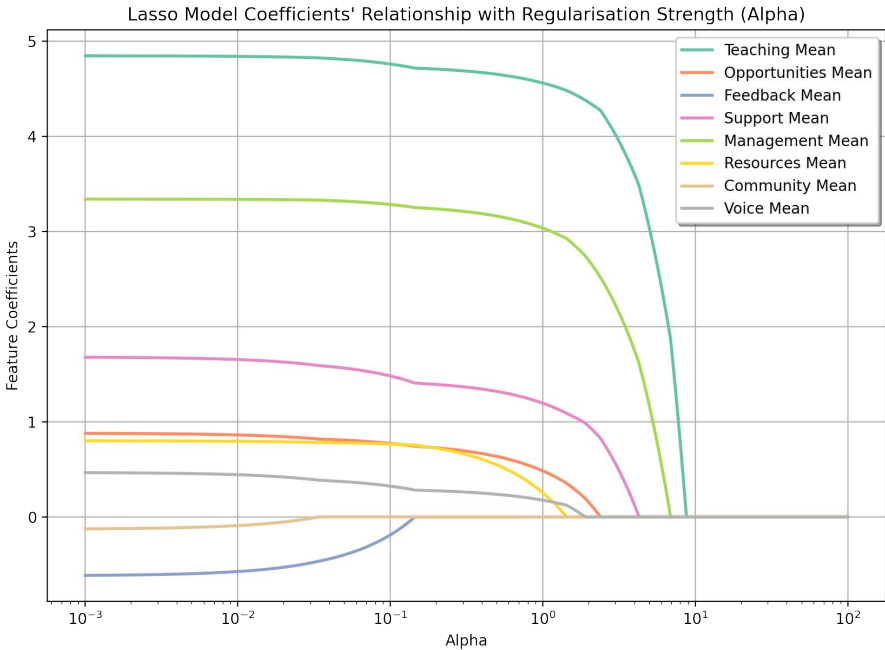
I have received sufficient advice and guidance in relation to my course



Good advice was available when I needed to make study choices on my course

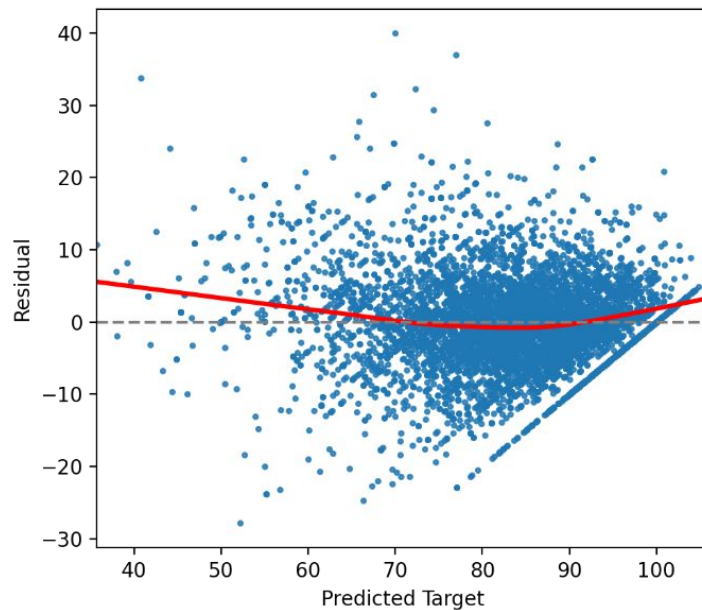
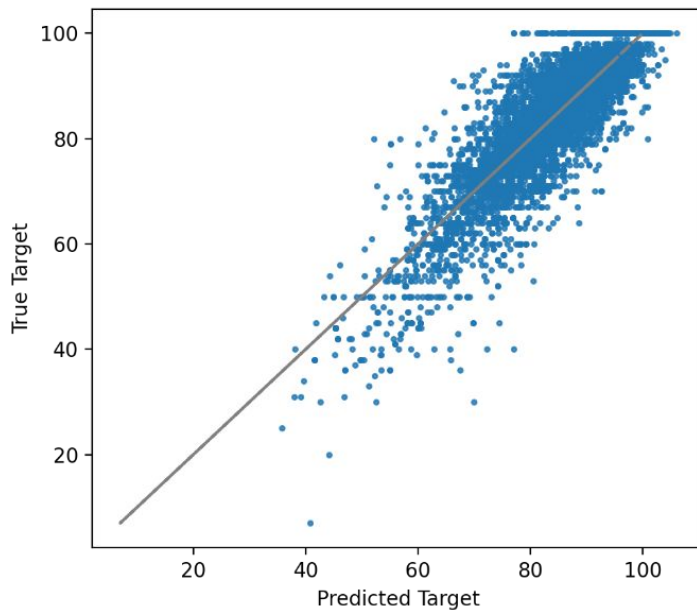
Results

Lasso Path



Results

Linear Regression Model (all 26)



Test Set Results: $R^2=0.7366$ MAE=4.26

Conclusions & Future Work

- 1 Courses should focus on improving their Teaching, Management and Support
 - 2 Due to the nature of the data, the model best performs on higher value targets
 - 3 The overall satisfaction score may require other features to be better modeled
-
- 1 Combine location data of universities with regional statistics
 - 2 Explore the effect of polynomial features