Slide 1

- Say hello, thank for the opportunity to speak to them. Indicate the presentation is designed to last about 20 minutes.
- Introduce the general problem in terms of the stolen diamonds and the need to validate the price.
- Indicate that the task will be better done if we consider the way lapidarists usually value the diamonds. The basic concept is the 4C's...

Slide 2

- A chemically pure and structurally perfect diamond has no hue, like a drop of pure water, and consequently, a higher value.
- GIA universalized the D-to-Z Color Grading Scale. D is the best

Slide 3

- Natural diamonds are the result of carbon exposed to tremendous heat and pressure deep in the earth. This process can result in a variety of internal characteristics called ‘inclusions’ and external characteristics called ‘blemishes.’
- GIA has also a scale

Slide 4

- Diamond cut means how well a diamond’s facets interact with light.
- Of all the diamond 4Cs, it is the most complex and technically difficult to analyze. To determine the cut grade of the standard round brilliant diamond – the shape that dominates the majority of diamond jewelry – GIA calculates the proportions of those facets (on image) that influence the diamond’s face-up appearance.

Slide 5

- To put it simply, diamond carat weight measures how much a diamond weighs.
- A metric "carat" is defined as 200 milligrams. Each carat is subdivided into 100 points

Slide 6

- Frame the stolen diamonds in terms of those features via a visualization. They are just 10 diamonds and in just one category we have two diamonds.
- Talk about the other characteristics we have: Depth, Table, x, y, z, latitude, longitude
- This representation leads to the idea that there are classes of comparable diamonds

Slide 7

- We have the same visualization for our big database.
- There is a pattern related to the color: the database has high quality color diamonds. In cut we don't have the worst nor the best category.
- We have, then, the manual solution: already on the slide.

Slide 8

- Results of manual solution. It is completely instance based, and is influenced by noise.
- But it's a good start

Slide 9

- Data science enters
- Indicate linear trend
- Indicate differences among colors, cuts and clarities
- Indicate linear regression is a good starting point, due to linear trend and categories.

Slide 10

- Since we want to validate is better to give ranges to point estimates.
- If point estimation, how do we know if the claimed price is close enough?
- Some information is redundant: depth and table are coded into cut, x,y and z seem to be lenghts that, when multiplied, give something with strong correlation to Carat.
- We can eliminate that redundant information for a first model.

Slide 11

- Difference between limits of prediction interval and interval for the mean. The interval for the mean is better estimated.
- We wll consider mean prices to be fair prices!

Slide 12

- The linear model captures all the known relations among different values of the 4C

Slide 13

- Results. Each interval is 95% for the mean.
- Closing advice (under the analysis so far): if the claim is under 24,990 and you agree that mean value is fair value, take it as good.

- Explain median
- Ask if they think median is the best concept for fairness. If that's the case, we are overestimating the value.
- How much? I'm not so sure. I didn't have enough time to perform the analysis!