



DATA 106 Notes

Process for basic Data Science

Goal No.	Skills	Details	Property	Property 1
<u>Goal 1</u>	Picking a topic	Identify the Institution – you will be required to collect real data, so be specific about the institution		
<u>Goal 2</u>	Identify a problem	Choose a relevant problem that the institution might be facing. One where you can find some data that may be able to help you answer the questions.		

Goal No.	Skills	Details	Property	Property 1
<u>Goal 3</u>	<div>Create Hypotheses</div> <div>Finding relevant data</div>	<p>You will come up with at least 3 hypotheses. At the initial stage, ask as many questions as possible then, narrow the list down. Consider the dataset you collected. You can only answer questions and test hypotheses for which you have relevant variables in your dataset.</p>		
<u>Goal 4</u>	<div>Data Cleaning</div>	<p>Cleaning the data is an essential step in the data analysis process. Without this step your results might be useless and might be measuring trends that are as a result of 'dirty' data</p>		

Goal No.	Skills	Details	Property	Property 1
<u>Goal 5</u>	Exploratory Data Analysis (EDA)	look the relevant variables in the dataset (for example in pairs dependent on the questions you are asking). After this step, you should be able to provide a qualitative and quantitative description of the dataset. (This might give you an insight into the answers to your questions and will help you check if assumptions are met)		
<u>Goal 6</u>	Fit models	use any model/method to answer your questions		
<u>Goal 7</u>	communicate your results	You must communicate your results in a non-technical way: a.) so, everyone in the audience understands the story (comprehensible and compelling), and b.) to keep the attention of the audience create a presentation that allows the audience to visualize the data and understand the results.		
<u>Untitled</u>				
<u>Goal 1:</u>				
<u>Goal 2:</u>				