Here are some Questions/Answers students had on the data given to them for their case study.

1) In the dataset, is the distance from home in miles or kilos?

We don't have that information (however we do know whether its high or low)

2) In the dataset: what is the definition of pay rates: Hourly, Daily & Monthly.  These values to not seem to relate to each other or the Monthly Salary (which is different than Monthly Rate).

We don't have that information (however we do know whether they are high or low). They may or may not relate to each other or the monthly salary (this is for the student to infer and decide whether there’s any correlation or whether this is a useful feature for attrition)

3) In the dataset: we do see that Job Levels go from 1-5 and assume that 1 may symbolize a lower level employee, but this is not defined.  Though this level does have a positive linear relationship with Monthly Income, it does not seem to correlate well with the Job Titles. in other words someone with a Director can be a 2-5, and manager a 3-5.

Yes we can assume 1 is a lower job level than 5.

4) In the dataset, does overtime mean Hourly vs. Salaried worker?

We can assume that people with overtime are non-exempt employees.

5) In the dataset, Performance Ratings are only 3 & 4, is there a mistake?  Unless a corrupted system, hard to imagine ratings consistantly high, even as 2 still means "good".

It is self-reported data, think about why the employees may only answer 3 and 4

No this is the only data we have, there is no mistake. Think about why employees were only answering 3 or 4.

6) In the dataset, does Training times mean: hours, weeks, or instances and over what period

Training times last year means number of training sessions attended by the employee.

7)  hourly rate, monthly rate ,daily rate, monthly income - these fields should have been related, but when we look at the data , there seems to be no connection. For ex - If we do daily rate/24, does not match hourly rate or id we do monthly rate/daily rate i would expect to get approximate days worked in a month which i would not expect to cross 31. But there are values > 200. Can we get definitions for these fields so we could apply correct logic for the values.

We don't have that information (however we do know whether they are high or low). They may or may not relate to each other or the monthly salary (this is for the student to infer and decide whether there is any correlation or whether this is a useful feature for attrition)

8) Standard Hours = 80 , mostly work hours are 40 per week. Each record shows 80 hours, looks like it is standard hours per bi-weekly pay check.  Is our assumption valid?

Yes

9) Num companies worked - is it num companies prior to current company? We some some of the records have a '0' but then total years of work <> 0, so looks like this field represents number of companies worked prior to current company

Yes it is prior.

11) Management positions - This is w.r.t task 3-d, By management positions are we expecting Executive positions? There is no direct field and we think we go either by job level ( > 3 or only 5 ? ) or by the job roles ?

That depends on how you want to conceptualize and operationalize Management positions. Is a manager a CEO? What does research/best-practice suggest on what is a management position?

12) Life Satisfaction - For task 4-d , there is no direct field for Life Satisfaction. do we need to create our own set of variables to analyze this feature ? So it would be based on our assumption ?

Yes, you have to figure out how you will create that given the data you have. Once again what is your concept and how will you operationalize it?