1.数据的生命周期管理包括6个阶段： 数据采集，数据存储，数据处理，数据传输

，数据交换，数据销毁。

2. 数据采集是指从传感器和其它待测设备等模拟和数字被测单元中自动采集非电量或者电量信号,送到上位机中进行分析，处理。数据库采集，系统日志采集，网络数据采集，感知设备采集是数据采集一些的方法。

3. 数据管理是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程。大数据管理技术数据规模量较大。传统数据技术主要是利用现有存在关系性数据库中的数据，对这些数据进行分析、处理，找到一些关联，并利用数据关联性创造价值。这些数据的规模相对较小，可以利用数据库的分析工具处理。而大数据的数据量非常大，不可能利用数据库分析工具分析。同时，大数据管理技术具有非结构化数据的特点。传统数据主要在关系性数据库中分析，而大数据可以处理图像、声音、文件等非结构化数据。

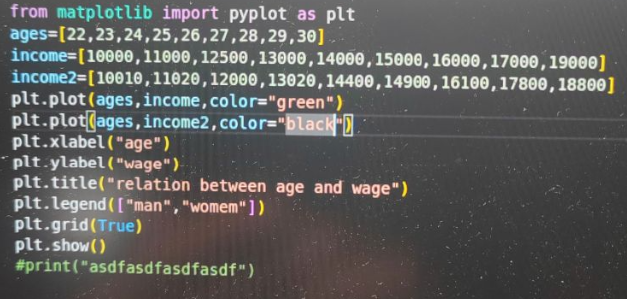
4.大数据的计算模式可以分为4类：批处理计算，流计算，图计算，查询分析计算。

5. 数据分析是指用适当的 统计分析 方法对收集来的大量数据进行分析，将它们加以汇总和理解并消化，以求最大化地开发数据的功能，发挥数据的作用。数据分析的方法有描述统计、假设检验、信度分析、列联表分析、相关分析、方差分析、回归分析、主成分分析、簇分析、聚类分析、决策树、KNN、SVM、随机森林、贝叶斯等

6. 数据可视化是一种将数据转换成图形化表达的技术，通过图表、地图、图形等形式来呈现数据，以便更好地理解数据的含义和趋势。 数据可视化已经成为大数据时代的重要工具之一，因为它可以帮助人们更好地处理和理解大量的数据，以及在业务和决策中发现数据背后的价值。

7.使用matplotlib绘制折线图:绘制了一个男女性的收入虽年龄变化的图像（虚构）。

用ages存放年龄作为横坐标，income和income2作为男女的收入数据。并调用plt.plot函数将横纵坐标作为参数传入。用plt.xlabel,plt.ylabel为坐标轴取名。plt.title为标题取名。plt.legend表示出曲线的含义。plt.gird()显示表格。最后调用plt.show战士折线图。代码和效果图如下:





8.绘制饼图:下面绘制一个普通家庭支出占比的饼图。用channels表示支出的渠道，sale表示支出的金额，explodes表示强调，因为食物占比最多，这里将食物强调出来。最后用plt.pie绘制出饼图并用plt.title()写出图的名字。最后在调用plt.show()将图展示出来。

