大家好！

我的报告主题是“麦当劳菜单的营养成分可视化与分析”。之所以选择这个主题，主要是因为现代人越来越注重身材管理，却又不能辜负美食，因此如何在吃的同时控制好营养均衡就变成了一个问题。

数据来源是kaggle网站，数据的形式是一个csv文件。

拿到数据后首先读取文件头信息，可以看到这个文件包含种类、食物、分量以及卡路里、脂肪等一共21种营养成分的数据，并且没有空值，因此没有进行额外的数据处理。

接着是数据的可视化。

这是麦当劳所提供的不同类别的食物的数量的柱状图和饼状图。可以看到，麦当劳提供了相当丰富的咖啡茶类，高达95种；其次是早餐，42种；然后是炸鸡炸鱼类、饮料类、冰沙奶昔类27，28种；牛肉猪肉类和小事均是十几种；甜点和沙拉类提供的选择是最少的，分别是7种和6种。

然后对各营养成分进行可视化和分析。

刚才提到数据中一共包含21中营养成分的信息，如果要一一进行可视化和分析，过程会比较复杂冗长，所以我决定先对各营养成分两两之间的相关性进行分析，这里采用到seaborn库中的热力图。

由于属性较多，上面这种热力图看起来还是不够清晰，有两个点可以帮助简化：

1. 可以明显发现矩阵中的数据有重复，比如'Total Fat'和'Total Fat (% Daily Value)'等，这很容易理解，因为后者是在前者的基础上与每日所需量的比值，因此在分析各营养成分两两之间的关系时不需将类似'Total Fat (% Daily Value)'的数据加入其中；

2. 这是一个对称矩阵，只需取一半就可以了。

右图是简化后的热力图。可以看出卡路里、脂肪、钠、蛋白质和铁（%每日所需）之间的相关性较大。

为了更加直观地分析上述营养成分之间的关系，这里采用seaborn库中的pairplot图来进行展示。pairplot中pair是成对的意思，pairplot主要展现的是变量两两之间的关系（线性或非线性，有无较为明显的相关关系）。

卡路里和脂肪之间存着明显的线性关系，且和食物类别没有太大关系；而卡路里、脂肪、蛋白质和钠之间也存在着明显的线性关系，但是和食物类别有着较大关系。

因此减肥时期最应注重的是脂肪的摄入，并且从非肉、蛋类的食物中摄入蛋白质容易导致卡路里超标。

接下来对一些营养成分进行单独的分析。

首先是我们最关注的卡路里，也就是热量。

这张图描述了麦当劳食物总体的热量分布情况。可以看到麦当劳每份食物的热量主要集中在250-500卡路里。同时计算了一下热量的均值是368.26千卡，中值是340千卡，热量较高，对于减脂或保持身材的女孩子来说一份食物的热量就已经达到了一餐的限制。

再来看一下各种类的食物的热量分布情况，为了同时展示出各种类中食物热量的分布情况以及各种类之间热量的比较情况，我采用的是seaborn库中的分簇散点图。

可以看到，早餐、牛肉猪肉、炸鸡炸鱼类食物每一份的热量主要分布在250-750kcal之间，集中在500kcal附近，其中早餐和炸鸡中有几个离群点，是由于超额的分量造成的，可以忽视；而沙拉、小食、甜点的热量在250kcal左右；饮料类在0-250kcal之间；咖啡和茶类在0-750kcal之间，集中在250kcal上下；冰沙奶昔类是饮品中热量最高的一类，在250-800kcal之间浮动。

因此，如果一定要喝点什么的话，建议普通饮料和咖啡茶类，和奶昔类相比更不容易发胖。

脂肪数据的分布情况与卡路里基本一致，与之前在热力图中得出的结论相符。

饱和脂肪酸是一种不太好的成分，来自于大部分动物油脂和少部分植物油中，图中可以看到早餐、牛肉猪肉以及奶昔类含有饱和脂肪酸的含量偏高。猜测前两者是动物油脂含量偏高的原因，而奶昔应该是牛奶导致的。比较意外的是炸鸡炸鱼和小食的饱和脂肪酸含量并不是很高。

反式脂肪酸是一种非常不健康的成分，应当避免摄入。可以看到牛肉猪肉类含有反式脂肪酸水平较高，应当少吃；少部分早餐、炸鸡炸鱼、咖啡茶类和奶昔也含有一定反式脂肪酸。

胆固醇主要由人体自身合成，食物摄入是补充。麦当劳食物中胆固醇含量从总体来说并不高，但早餐中胆固醇含量偏高。

一般说来，对碳水化合物没有特定的饮食要求。主要是应该从碳水化合物中获得合理比例的热量摄入。奶昔碳水偏高，不建议作为饮品的选择。

最后是一些其他的营养成分，钙、铁、维生素A、维生素C都是重要的微量元素。因为这些元素与食物种类的关系不大，所以使用交互式散点图进行可视化，因此在jupyter notebook上展示。