**INFO 103 – Project description**

Group member names:

Group member IDs:

Assigned: 10/23/19; Due: 11/10/19

Instructions: Groups may have a maximum of **four** and a minimum of **two** people. Turn in one document for your whole group. （presentation）

1. **(5 pts)** Your project should fit into one of three categories. Please select one, below.
   1. **Data product or project development.** Study the development of an existing or hypothetical data product or project, focusing on value and marketability.

**数据产品开发：**“数据产品”在数据科学中具有特殊的含义——基于数据开发的产品的统称。数据产品开发是数据科学的主要研究使命之一，也是数据科学区别于其他科学的重要区别。与传统产品开发不同的是，数据产品开发具有以数据为中心、多样性、层次性和增值性等特征。数据产品开发能力也是数据科学家的主要竞争力之源。因此，数据科学的学习目的之一是提升自己的数据产品开发能力。（<https://www.sohu.com/a/224047326_774914>）

农业-交通-配送一体化

* 1. **Data collection and curation.** Study an existing or hypothetical data source, focusing on its challenges in curation, storage, and collection.
  2. **Data technology and processing.** Study an existing or hypothetical processing methodology or big data technology, focusing on problems and applications.

1. **(15 pts) Who’s on this project’s team?** Whether existing or hypothetical, what team member specialties did or would this specific project require? Identify three or more **roles** from the data science spectrum, and discuss how the skills of the individuals filling these roles support the needs of this data science project. **Note: it may help to complete task 3 before completing this task!**

**Use more Week 1 matrices**

**2.（15分）谁是该项目团队的成员？ 无论是现有的还是假设的，该团队成员需要做什么专业知识？ 从数据科学领域确定三个或更多角色，并讨论填补这些角色的个人技能如何满足该数据科学项目的需求。 注意：这可能有助于完成任务3，然后再完成此任务！**

**使用更多的第1周矩阵**

数据收集、数据处理、数据储存对应的科学家

1. **(25 pts) Identify your specific topic of study, and provide a summary, below.** Now that you have selected your project category, you must select and describe the specific topic on which you will focus. Using complete sentences, this should include a clear description of any **product**, **data**, **market**, **technology**, or **processing method**. In addition, this section should include a discussion of the topic’s **impact** and **use**. **This summary should be 1–2 pages in length, single spaced at font size 12.**

**3.（25分）确定您的特定学习主题，并在下面提供摘要。 现在您已经选择了项目类别，您必须选择并描述您 将关注的特定主题。 使用完整的句子，应包括任何产品，数据，市场，技术或加工方法的清晰描述。 此外， 本节还应讨论该主题的影响和使用。 此摘要的长度应为1-2页，单行间距为12号字体。**

[[1]](#endnote-1)Topic of study:

创造一个数据综合管理系统:管理农场和物流， 在保证食品是健康、安全的前提下，使农场主和农产品加工商能保证农产品（蔬菜、猪肉、牛奶等）快速地从农场送到餐桌上。

Description of:

Product:

Data: 传感器收集到的农产品的生长数据、物流运输的实时数据（交通情况、最优路线、API---交通管制系统实时检测的数据）

Market: 单个用户、餐厅、超级生鲜市场（沃尔玛、盒马···）

Technology: LoRa（远距离无线电技术） & IoT（物联网技术）& Sensor（传感器）

1

市场：单个客户、餐厅、超级市场（沃尔玛、合马）

算法估计周围的人口密度等来计算需求量

数据：根据之前的各个商品的购买量来预测当天的需求量，然后配送。

在一开始，我们的农场就要确定目标产量（根据当地城市的具体情况 人口 习惯啥的，这些信息的收集就需要大数据，调查啥的） 因为只有这样才能定量比较垂直和传统哪个方案更优秀。（这个是展开选址前的第一步）

>>>

新鲜果蔬是居民的必需品，包括水果、蔬菜等，对居民的生活有着重要的影响。传统生鲜行业采用“产地—各级批发市场—菜市场”的模式，中间环节多，运输效率低，终端价格高。

我们的这个项目设计的data product 的市场主要有：个人，餐厅，超级市场，生鲜市场。

首先需要根据收集和统计的数据进行市场调查，市场需求调查：包括市场的商品需求种类、市场商品需求量、及需求时间。即了解消费者在何时何地需要的商品种类和数量。市场环境调查：了解市场购物习惯。那样就可以控制农场的目标产量，起到成本控制的作用。

通过计算各个地区的市场需求，我们将农场选在最适宜的地方，目的是为了使物流更加便捷，降低仓储和物流的成本。

Market

Fresh fruits and vegetables are a necessity for residents and have an important impact on the lives of residents. The traditional fresh-keeping industry adopts the model of “place of origin – wholesale market at all levels – vegetable market”, with many intermediate links, low transportation efficiency, and high terminal prices.

The market for the data product designed by our project mainly includes individuals, restaurants, supermarkets, and fresh markets.

First, we need to do market surveys based on collected and statistical data. Market demand surveys include the types of commodity demand in the market, the demand for market commodities, and the demand time. That is to understand the types and quantities of goods that consumers need when and where. That way we can control the farm's target production and the cost of the farm.

Then, by calculating the market demand in each region, we select the farm in the most suitable place, in order to make logistics transport more convenient and reduce the cost of warehousing and logistics transport.

1. **(25 pts) Support your topic with reference materials and briefly summarize what each piece of reference material tells you about your topic.** These materials should relate to key components of your topic, and might include **academic papers**, describing related methods or data sets; **news articles**, indicating market demand, changing use, or ethical concerns; **blog posts**, describing products, updates, and opinions; **patents**, describing system details; **APIs**, describing data and software access; or **any documentation** detailing structure, licensing, and use. Note that all projects should have substantial reference material, **regardless of topics being existing or hypothetical.** For example, a hypothetical project might focus on references for market potential, competition, and predecessor methods, while an existing project might focus on its history, system descriptions, product updates, and ethical impacts.

4**.（25分）用参考资料支持您的主题，并简要总结每篇参考资料告诉您的主题。 这些材料应与您主题的关键 组成部分相关，并且可能包括学术论文，描述相关方法或数据集； 指出市场需求，用途变化或道德问题的新 闻文章； 博客文章，描述产品，更新和意见； 专利，描述系统细节； API，描述数据和软件访问； 或任何详 细说明结构，许可和使用的文档。 请注意，无论存在的主题还是假设的主题，所有项目均应具有大量参考资 料。 例如，一个假设的项目可能专注于市场潜力，竞争和先前方法的参考，而现有的项目可能专注于其历 史，系统描述，产品更新和道德影响。**

**Reference [1]**

[《一篇完整的市场调查方案、计划、策划书包含哪些内容》](https://wenku.baidu.com/view/350318c5777f5acfa1c7aa00b52acfc789eb9ff1.html)

Original text cited：

“市场需求调查市场商品需求调查主要包括市场商品需求量、需求结构和需求时间的调查。即了解消费者在时何地需要什么，需要多少。市场商品需求量主要取决于社会购买力水平。调查市场商品需求量主要是调查社会购买力，是一种有支付能力的消费需求。对企业来说，调查市场需求量，不仅要了解企业所在地区的需求总量、已满足的需求量和潜在的需求量，而且还必须了解本企业的市场销售量在市场商品需求量中所占的比重，即本企业销售的市场占有率，以及开拓地区市场的可能性。”

Topic:

Refer to the market survey method to collect more objective data for market research from various aspects.

**Reference [2]**

[《我国生鲜市场调研报告》](https://wenku.baidu.com/view/aacd0b61951ea76e58fafab069dc5022aaea469c.html)

Original text cited：

“随着“互联+农业”趋势不断深化，生鲜电商保持着高速发展的态势，正日益成为电商行业的突破性增长点。巨大的市场前景吸引了众多生鲜电商的积极入市：京东7000万美元领投天天果园；阿里巴巴投资易果；顺丰跨界开办顺丰优选；苏宁超市推出“苏鲜生”；亚马逊公布了筹备已久的生鲜馆，包括21cake、都乐、獐子岛等；联想佳沃推出金艳果猕猴桃，“柳桃”与“橙”、“潘苹果”一起掀起水果络营销的浪潮。   新事物的发展不可能是一蹴而就的，行业的成长阶段必然要经历艰难的探索时期。生鲜电商普遍存在盈利难的问题，截至20XX年7月份，国内4000多家生鲜电商，仅40家实现盈利，95%的生鲜电商企业仍处于不同程度的亏损，其中包括京东、天猫、顺丰优选等生鲜电商巨头。大型综合电商们有雄厚的资金支持，短期亏损尚能维持，但许多中小型生鲜电商由于资金链断裂而被迫退出市场，曾经风光一时的优菜、美味七七、菜管家等垂直电商的失败，显示出生鲜电商面临的严峻形势。生鲜电商虽然道路曲折，但其发展前景仍被业内所看好，资本的青睐为生鲜电商的发展注入了新的活力。20XX年资本“寒冬”的大背景下，中粮我买完成了C轮亿美元融资，创下行业新高，本来生活、天天果园、多点等其它多家公司也完成了新一轮融资。传统农批市场转型电商的实力强大，成为生鲜电商发展的新增长点。面对生鲜电商的迅猛发展给传统生鲜行业带来的>中击，许多农批市场积极转型，借力互联信息化大趋势，开展农产品电子交易、电子拍卖和期货交易等业务。例如，寿光农产品物流园建立起蔬菜电子拍卖中心，日交易量达100多万斤，成功实践了“未收先卖，未种先卖”的订单农业模式。”

Topic:

Investigate and study the current situation of China's fresh market.

**Reference [3]**

[基于LoRa的智能农业系统设计与实现.pdf](参考资料/基于LoRa的智能农业系统设计与实现.pdf)

[What is LoRa?](https://www.semtech.com/lora/what-is-lora)

Original text cited：

“设计一套基于LoRa技术的智能农业系统,能够实现实时准确地采集农作物生长环境中各种参数,并能 实时精准地控制大棚中的设备,让大棚中的生长环境保持 在一个平衡的状态,从而满足农作物生长的需求。从现实 意义上看,本系统可以提高农业生产效率与农业产量,同时改善农业安全以及粮食安全。

智能农业系统从整体上由 四部分构成,分别是:用户访问平台、农业云平台服务器、 LoRa基站和终端节点。其中用户访问平台指的是 PC 端和 手机客户端,通过它们来监控大棚内的环境情况。农业云 平台服务器负责对整个系统的数据进行处理,并且在其数据库内进行数据保存,LoRa基站负责对终端节点和农业云 平台服务器之间数据的交互,终端节点包括采集节点和控制节点。

温湿度采集节点、光照采集节点、CO2 采 集节点等称为采集节点,同时控制节点包括风扇控制节点、 卷帘控制节点、加热控制节点等,另外一个 LoRa基站负责 多个农业大棚的终端节点,一个农业云平台服务器负责多个LoRa基站。在本系统的总体设计中支持上与下双向数据 传输。从下至上时,首先通过采集节点采集到大棚内温湿 度、光照强度、CO2 浓度等信息的数据,经 LoRa 网络自组 网方式传输到LoRa基站,然后经过3G/4G/有线宽带网络 上传到农业云平台服务器,实现与用户访问平台的对接。 从上至下时,用户访问平台首先通过互联网发送指令给农 业云平台服务器,然后经由 3 G/4 G/有线宽带网络发送给 LoRa基站,最后再经过 LoRa 无线网络发给控制节点,通 过控制风扇、卷帘、加热器的开关来调节大棚内的环境, 从而实现了对控制设备的控制。“

Topic:

1. Learn about LoRa (short for long range).

2. Research on integrated sensors: Use sensors to collect data on soil fertility, soil moisture, trace elements, light and other growth environments and to monitor plant growth status.

**Reference [4]**

[“Drones and Neural Networking Used to Identify Citrus Trees from Above”，Stefan Tasevski](https://dronebelow.com/2018/11/27/drones-and-neural-networking-used-to-identify-citrus-trees-from-above/)

Original text cited：

“ Nowadays, remote sensing is one of the processes used in precision agriculture. The use of drones or unmanned aerial vehicles (UAVs) in this field is revolutionizing precision agriculture workflows for measurement of crop conditions as well as yields over the growing season. Aside from this, practices in the monitoring of trees for growth, individual fruit production, monitoring the weeds and others are found to be very useful for long-term farm management.

UAV imagery collection for agricultural applications is increasing globally and more of these individual cases are needed to develop more standard workflows that will help field and research managers deal with large volumes of high resolution imagery.

Methods to delineate, enumerate and monitor individual trees in an agricultural setting from high resolution optical imagery are required for efficient and precise crop management. Monitoring of individual trees for growth, fruit production and pest and disease occurrence remains a high research and operational priority and the delineation of each tree using automated methods as an alternative to manual delineation will be useful for future long-term crop management.

”

Topic:

1. UAVs represent a low-cost method for image acquisition with successful and promising usage in tree identification analysis, particularly in agricultural settings.

2. Research on image recognition (analysis of plant growth and results using images).

3. Research on the application of small civilian drone technology.

**Reference [5]**

[The\_Effects\_of\_Cold\_Chain\_Logistics\_and.pdf](参考资料/The_Effects_of_Cold_Chain_Logistics_and（冷链技术对全球的影响）.pdf)

Original text cited：

“While Globalization has made the relative distance between two regions of the

world vastly smaller, the physical separation of these same regions is still a very

important reality. Phone calls, emails and videos can be transmitted in fractions of

seconds to all corners of the globe, but physical objects such as a bushel of grapes, a drug

or a bodily organ cannot. It takes time and coordination to efficiently move a shipment

and every delay can cost money and in some cases may even cost lives. To ensure that

cargo does not become damaged or compromised throughout this process, businesses in

the pharmaceutical, medical and food industries are relying more and more on the cold

chain technology.

Current and Future Innovations

As the pharmaceutical and food industry companies that rely on cold chain

transportation for their economic livelihoods move into the future, new technological

advancements and inter-industry cooperation seminars aimed at strengthening every link

in the supply chain will continue to emerge. One growing technological improvement to

the supply chain is radio frequency identification (Refrigerated Transportation 2006) or

the use of identification tags that provide up to the minute reports of where a shipment is

located anywhere in the world (Murphy-Hoye et al 46). The use of a tracking device such

as these offers the prospect of being able to provide total accountability and service

control while a shipment is moving through the supply chain. Any delays or location

questions will be able to be answered immediately, making logistical pre-planning much

easier. RFID technology also reduces the amount of labor hours needed for tasks

associated with accounting for and recording stock piles of inventory (Murphy-Hoye et al

46). The tracking system also offers better protection against potential counterfeit drugs

infiltrating the pharmaceutical market (Basta 1). ”

Topic:

Research on urban transportation and cold chain distribution.

**Reference [6]**

[Design\_and\_fabrication\_of\_Windchill\_for.pdf](file:///C:\Users\47413\Desktop\数据科学导论\lab\参考资料\Design_and_fabrication_of_Windchill_for（低成本蔬菜储存）.pdf)

Original text cited：

“Windchill is the lowering of body temperature due to the passing-flow of lower-temperature air. In the windchill food preservation system, only a fan sucked the air into the chamber through aluminum tubing. The tube was buried into the ground where the temperature is less than the atmosphere. When the air was passed through the tubing it was cooled by exchanging the heat with soil via aluminum tubing and also by the expansion of the volume of air when entering into the chamber. Temperature and relative humidity (rh) are two most important parameters that affect the shelf-life of vegetables. It was found that low-temperature storage protects the quality of vegetables like texture, nutrition, aroma, and flavor (Paull, 1999). The freshness quality of vegetables has an influence on the market price of the vegetables. Low temperature and high humidity also reduce the loss of moisture which results in a low weight loss of vegetables. So it is needed to control the temperature and rh to increase the shelf-life of vegetables and maintain the freshness quality.

A windchill vegetable storage chamber was designed and fabricated in the FPM departmental workshop. This study was adopted on the principle that some insects keep their living place comfortable under the ground and on the principle of ideal gas. As the temperature was low and rh was high inside the chamber from the outside, vegetables kept inside the chamber was found fresher and weight loss was less as compared with those kept outside for 3-5 days. From the panel test result, it was assured that the vegetables were consumable after four days which kept inside the chamber. Based on this study, it can be concluded that this storage chamber can be used as temporary means of cost effective storage in farmers’ levels for several days.”

Topic:

Research on the storage of crops.

**Reference [7]**

[经济学人0831.pdf](参考资料/经济学人0831.pdf)

Original text cited：

“From the outside, it looks like a tall, metal-clad barn. But step in, through a large airlock designed to keep out the bugs, and a kaleidoscopic scene emerges. A central aisle is flanked by two pairs of towers. Each tower is stacked with a dozen or so trays on which are growing strawberries, kale, red lettuce and coriander. And each tray is bathed in vibrant light of diﬀerent colors, mostly hues of blue and magenta. Douglas Elder, who is in charge of this Artiﬁcial Eden, taps some instructions into an app on his mobile phone and, with a short whirr of machinery, a tray of lush, green basil slides out for his inspection.  Mr. Elder is a product manager for Intelligent Growth Solutions (IGS), a “vertical farming” company based at Inver Gowrie, near Dundee, in Scotland. Each of the nine-meter-high towers in the demonstration unit that he runs occupies barely 40 square meters. But by stacking the trays one on top of another an individual tower provides up to 350 square meters of growing area. Using his phone again, Mr. Elder changes the colors and brightness of the 1,000 light-emitting diodes (LEDs) strung out above each tray. The app can also control the temperature, humidity and ventilation, and the hydroponic system that supplies the plants, growing on various non-soil substrates, with water and nutrients. Armed with his trusty phone, Mr. Elder says he can run the farm almost single-handedly.”

Topic:

Research on vertical farms.

**Reference [8]**

[Natural Disaster Prevention](https://sites.google.com/site/naturaldisasterpreventiongin/prevention-prediction)

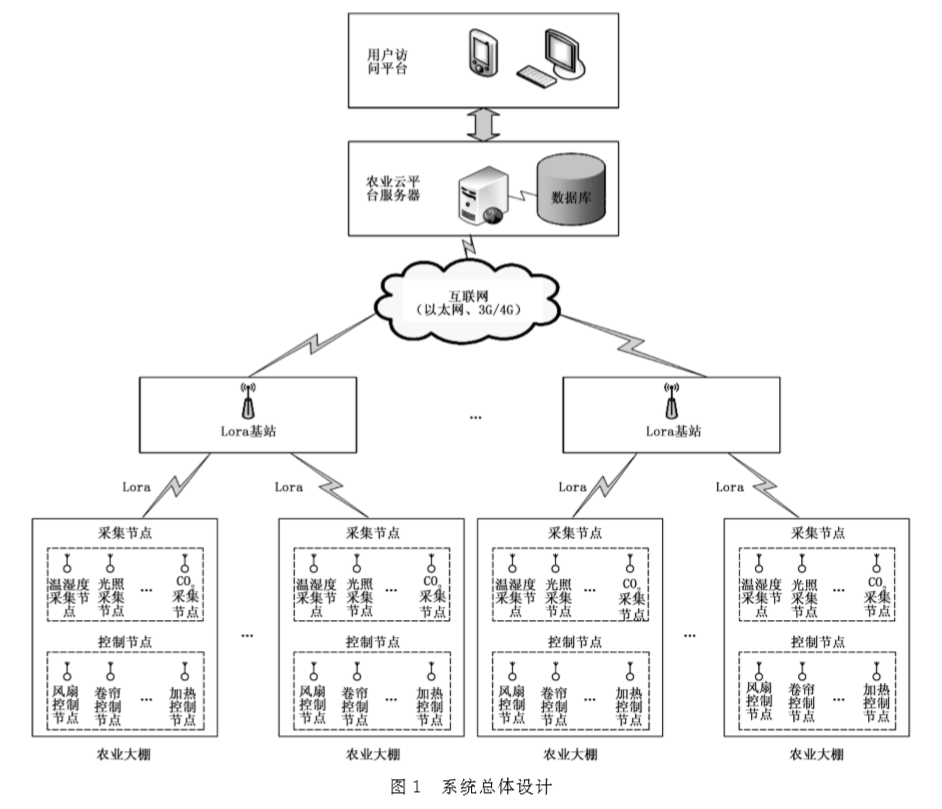
Topic: Natural disaster prevention, providing technical support for farm site selection.

1. **(30 pts) Identify how your topic intersects with the different areas of the data science life cycle.** Each stage of the data science life cycle is listed below. Using complete sentences, fill in how your topic crosses over into each area. If your topic generally does not cross over into an area, please indicate why this is the case in that stage’s area. Note that even if your topic does not specifically engage in a particular stage of the life cycle, **it may still have significant impacts to other to other data science projects.** For example, even though a **data collection and curation** may have a greater focus on **preparation**, there may be important implications in **acquisition** for others who want to use this data. Be sure to include these impacts.

**5.（30分）确定您的主题如何与数据科学生命周期的不同领域相交。 下面列出了数据科学生命周期的每个阶段。 使用完整的句子，填写主题如何跨越每个区域。 如果您的主题通常不涉及某个区域，请说明为什么在该阶段的区域是这种情况。 请注意，即使您的主题没有专门参与生命周期的特定阶段，也可能会对其他数据科学项目产生重大影响。 例如，即使数据收集和管理可能更侧重于准备工作，对于希望使用此数据的其他人来说，数据采集也可能具有重要意义。 确保包括这些影响。**

**Data acquisition:**

**数据采集：**



终端节点包括采集节点和控制节点，图一中，温湿度采集节点、光照采集节点、二氧化碳采集节点等称为采集节点。

**Data preparation:**

**数据准备：**

**Hypothesis and modeling:**

**假设和建模：**

**Evaluation & interpretation:**

**评价与解释：**

**Deployment:**

**部署：**

**Operations:**

**操作方式：**

**Optimization:**

**操作方式：**

1. [↑](#endnote-ref-1)