一、调查分析数据科学家常用方法、技术与工具：

常用方法

1. 数据收集与清洗：数据科学家首先需要收集大量的数据，并进行清洗和预处理，以去除噪声、处理缺失值和异常值等，确保数据的质量和可用性。

2. 数据探索与可视化：在数据清洗完成后，数据科学家会进行数据探索和可视化分析，使用统计方法和可视化工具来了解数据的分布、关联性和趋势等，从而发现数据中的规律和模式。

3. 数据建模与预测：数据科学家利用各种建模技术来构建预测模型，以预测未来的趋势和结果。常用的建模方法包括线性回归、决策树、支持向量机等。他们还会使用机器学习和深度学习算法来训练模型，并通过交叉验证和调参等技术来提高模型的性能。

4. 数据挖掘与特征工程：数据挖掘涉及从大量数据中提取模式和知识，而特征工程则是构建和选择特征的过程，这对于机器学习模型的性能至关重要。

常用技术

1. 回归分析：用于估计变量之间的关系，预测连续型数据。

2. 聚类分析：将数据集分成由类似的对象组成的多个组或“簇”。

3. 决策树/决策规则：一种监督学习的方法，可以用于分类和回归。

4. 时间序列/序列分析：用于分析时间序列数据，以识别趋势和周期性。

5. 异常/偏差检测：识别数据中的异常或不寻常的模式。

6. 集成方法：结合多个模型以提高预测性能。

7. 支持向量机（SVM）：一种强大的分类技术，也可以用于回归分析。

常用工具

1. Python/R：数据科学家最常用的工具，在机器学习领域有非常丰富的库资源，两者都是开源并且免费的。

2. Apache Spark：一个提供实时数据分析的可扩展计算引擎，具有极快的数据处理能力。

3. RapidMiner：一个功能强大的数据科学工具，提供了可视化的工作流程设计，消除了繁琐的编码需求。

4. Neo4j Graph Data Science：一种解决方案，用于分析数据之间的复杂关系，发现隐藏的联系。

5. Jupyter Notebooks：广泛用于Python编程，支持多种语言，如Julia，R等。

6. SAS：一个模块化、集成化的大型应用软件系统，功能包括数据访问、数据储存及管理、应用开发、图形处理、数据分析等。

7. Weka：适用于数据科学从业人员，也是优秀的机器学习教学平台。

1. 调查分析近5年在数据科学与大数据领域出版的专著；调查关于数据挖掘领域的国内外主要学术期刊;调查数据挖掘领域顶级国际会议列表。

近5年在数据科学与大数据领域出版的专著：

1. 《R for Data Science》：由Hadley Wickham和Garrett Grolemund合著，以R语言为基础，介绍数据科学的基本原理和常用技术。

2. 《Python for Data Analysis》：作者是Wes McKinney，以Python语言为工具，教授数据分析的基本技能和方法。

3. 《Deep Learning》：作者是Ian Goodfellow、Yoshua Bengio和Aaron Courville，系统介绍深度学习的原理、方法和应用。

4. 《数据科学与大数据分析：发现、分析、可视化和呈现数据》：由EMC教育服务出版，涵盖数据科学家使用的各种活动、方法和工具。

5.《商业数据科学：您需要了解的有关数据挖掘和数据分析思维的知识》：由Foster Provost和Tom Fawcett撰写，介绍数据科学的基本原理。

6. 《深入浅出统计学：大脑友好指南》：作者是Dawn Griffiths，由奥莱利出版。

数据挖掘领域的国内外主要学术期刊

1. Data Mining and Knowledge Discovery：专注于从大型数据库中提取信息的技术期刊，发表原创研究论文和实践案例，提供重要领域和技术的调查与教程。

2. Big Data Mining and Analytics：通过挖掘和分析从各种应用中获得的大量数据来发现隐藏的模式、相关性、洞察和知识。

3. Data Science：跨学科期刊，涵盖从数据创建、挖掘、发现、管理到分析、预测、可视化等多个方面。

数据挖掘领域顶级国际会议列表

1. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)：数据挖掘领域的最高学术会议，由美国计算机学会ACM旗下数据挖掘和知识发现的专业组织SIGKDD主办。

2. IEEE International Conference on Data Mining (ICDM)：世界首屈一指的数据挖掘研究会议，提供国际论坛分享原创研究成果，交流和传播创新和实践经验。

3. SIAM International Conference on Data Mining (SDM)：由应用数学学会SIAM主办的数据挖掘会议，底子深厚，在计算机科学领域的影响力相对较小。