



# Introduction to the FDP Program

March 29 – April 18, 2021

Create your FDP Profile and receive the  
complete FDP Candidate Study Guide

Brought to you by



“Thank you for your interest in the FDP Charter. The present document is NOT the complete study guide. This document offers you an overview of what is available in the official study guide for the March 2021 exam. If, after reviewing this document, you conclude that the FDP Charter could be an essential step in your continued education and can help you gain a competitive edge in the data-driven finance industry, visit the FDP website, and create your FDP profile, and receive a copy of the complete study guide. Once you have decided to pursue the FDP Charter, register for the exam, and gain access to the sample exam and additional learning tools.”  
- *Hossein Kazemi*

# Outline of FDP Exam Study Guide Contents

<b>Introduction to the FDP Program</b> .....	3
<b>FDP Program: Online Requirements</b> .....	4
<b>FDP Examination</b> .....	6
<b>The FDP Curriculum: Outline</b> .....	7
<b>Other Study Tools and Resources</b> .....	8
<b>The FDP Curriculum: Reading List</b> .....	9
Topic 1: Introduction to Data Science & Big Data .....	9
Topic 2: Machine Learning: Introduction to Algorithms .....	9
Topic 3: Machine Learning: Regression, Support Vector Machine & Time Series Models .....	9
Topic 4: Machine Learning: Regularization, Regression Trees, Random Forest & Overfitting ....	10
Topic 5: Machine Learning: Classification & Clustering .....	10
Topic 6: Machine Learning: Performance Evaluation, Backtesting & False Discoveries .....	10
Topic 7: Data Mining & Machine Learning: Naïve Bayes & Text Mining .....	10
Topic 8: Big Data & Machine Learning: Ethical & Privacy Issues .....	11
Topic 9: Big Data & Machine Learning in the Financial Industry .....	11
<b>Example of Learning Objectives</b> .....	13
Topic 1: Introduction to Data Science & Big Data .....	13
<b>Action Words</b> .....	18

## Introduction to the FDP Program

The FDP Institute® was founded by the Chartered Alternative Investment Analyst Association® to create the FDP® charter. It is the only globally recognized professional designation in the area of financial data science, an increasingly important part of the financial services industry.

In recent years, the financial industry has been disrupted by the digital revolution. It is critical for industry practitioners to have a working knowledge of the increasingly important roles played by big data, machine learning, and artificial intelligence in the financial industry. The FDP Institute has designed this self-study program to provide the finance professional with an efficient path to learn about the essential aspects of financial data science. The FDP curriculum introduces candidates to the central concepts of machine learning and big data, including ethical and privacy issues, and their roles in various segments of the financial industry. Candidates will earn their FDP Charter once they pass the FDP exam and complete two short online classes, which can be done before or after the FDP exam.

The university faculty and industry practitioners who have helped create the FDP Charter program bring years of experience in the financial services industry. Consequently, the curriculum is consistent with recent advances in the applications of data science to the financial industry.

Passing the FDP examination is an important accomplishment and will require a significant amount of preparation. All candidates will need to study and become familiar with the FDP curriculum material to develop the knowledge and skills necessary to be successful on examination day.

The FDP Candidate Study Guide study guide is organized to facilitate quick learning and easy retention, and is available via [www.fdpinstitute.org](http://www.fdpinstitute.org). The study guide offers candidates a roadmap of learning. The exam has 9 topics, each topic is structured around learning objectives that define the content to be tested on the exam. The learning objectives are an important way for candidates to organize their study, as they form the basis for examination questions. All learning objectives reflect the content in the FDP curriculum, and all exam questions are written to address the learning objectives directly. A candidate who can meet all learning objectives in the study guide should be well prepared for the exam. For these reasons, we believe that the FDP Institute has built a rigorous program with high standards, while also maintaining an awareness of the value of candidates' time.

Candidates for the FDP Charter are required to complete both the FDP exam and the online classes. Since the FDP program is designed for finance professionals, it is assumed that candidates have an understanding of finance. This includes awareness of the roles and characteristics of various financial institutions and instruments as well as the financial models employed by these institutions to value the instruments and measure risk. These concepts are covered in CAIA®, CFA®, and FRM® exams, and dedicated undergraduate or graduate courses covering financial markets, investments, and risk management.

## FDP Program: Online Requirements

In order to receive the FDP Charter, candidates must successfully complete the following two components:

- **FDP Exam**
- **Two online classes covering the basics of Python or R programming, or the single class offered by Metis.**

Online classes are offered by three approved providers. The classes can be completed before or after the FDP exam. Depending on the candidate's background, the online classes are estimated to take 8-10 hours. No programming background is required to complete the online classes.

As of now, pre-selected online classes offered by the following organizations have been approved by the FDP Institute.

- **Datacamp:** <https://www.datacamp.com/>
- **Dataquest:** <https://www.dataquest.io/>
- **Metis:** <https://www.thisismetis.com/>

The list of online classes for each approved online provider appears on FDP Institute's website as well as in this document.

The approved online classes offered by Dataquest and Datacamp are available as soon as a candidate registers on their respective sites. The approved class offered through Metis is offered once a month throughout the year. All three providers offer limited free access to their classes. Candidates should take advantage of the limited free access to determine which platform's approach is more suited to their needs. Candidates cannot mix and match classes from different providers. Finally, the approved classes offered by the three platforms assume no prior knowledge of Python or R language or any specific computer programming language.

The Candidates' Handbook, which can be found on FDP's website, describes the procedure for sending proof of successful completion of the online classes to the FDP Institute. The following classes should be completed to satisfy the FDP Charter requirements.



#### DATA CAMP PYTHON TRACK

##### 1. Introduction to Python

In this Introduction to Python course, you'll learn about powerful ways to store and manipulate data, and helpful data science tools to begin conducting your own analyses.

<https://www.datacamp.com/courses/intro-to-python-for-data-science>

##### 2. Intermediate Python for Data Science

Learn to visualize real data with Matplotlib's functions and get acquainted with data structures such as the dictionary and the pandas DataFrame. After covering key concepts such as boolean logic, control flow, and loops in Python, you'll be ready to blend together everything you've learned to solve a case study using hacker statistics.

<https://www.datacamp.com/courses/intermediate-python-for-data-science>

#### DATA CAMP R TRACK

##### 1. Introduction to R

In Introduction to R, you will master the basics of this widely used open source language, including factors, lists, and data frames.

<https://www.datacamp.com/courses/free-introduction-to-r>

##### 2. Intermediate R

In this R training, you will learn about conditional statements, loops, functions to power your own R scripts, data structure manipulations, and times and dates. This course will allow you to take the next step in advancing your overall knowledge and capabilities while programming in R.

<https://www.datacamp.com/courses/intermediate-r>



#### DATAQUEST PYTHON TRACK

##### 1. Python for Data Science: Fundamentals

In this introductory course on Python for data science, you'll get an overview of the Python programming language and how you can use it for data science.

By the end of this course, you'll be able to: Understand the fundamentals of programming in Python, understand the fundamentals of data science.

Use Jupyter Notebook, and build a portfolio project.

<https://www.dataquest.io/course/python-for-data-science-fundamentals/>

##### 2. Python for Data Science: Intermediate

In this Python for Data Science Intermediate course, Dataquest covers some key techniques for working with the Python programming language for data science. By the end of this course, you'll be able to: Clean and analyze text data, Understand object-oriented programming in Python and Work with dates and times.

<https://www.dataquest.io/course/python-for-data-science-intermediate>

#### DATAQUEST R TRACK

##### 1. Introduction to Data Analysis in R

This is the first course in the DataQuest Data Analyst in R path, and in it, you'll be learning about the fundamentals of R. You'll learn to use variables, operators and write logical expressions. You'll also learn about the data analysis workflow in R.

<https://www.dataquest.io/course/introduction-to-data-analysis-in-r/>

##### 2. Data Structures in R

This is the second course in the Data Analyst in R path. In it, you'll build on the R programming skills you learned in the first course as you start to work with some of the most common data structures in R: vectors, matrices, lists, and dataframes.

<https://www.dataquest.io/course/data-structures-in-r/>



#### METIS

The single course offered by Metis is titled **Beginner Python and Math for Data Science**, and it consists of the following 6 topics covered during live sessions. Check the Metis website for upcoming classes. The 6 topics include: Python Basics, Python Advanced, Python Mathematical Libraries, Linear Algebra, Calculus and Probability, Statistics.

**APPROVED ONLINE CLASSES** The two (2) classes can be completed before or after the FDP exam. Depending on the candidate's background, it is estimated to take 8-10 hours to complete the online classes. No programming background is required to complete the online classes.

**SUBMIT YOUR CERTIFICATES** After you complete each class, you will receive a certificate from the provider. These Certificates of Completion must be submitted to [certificates@fdpinstitute.org](mailto:certificates@fdpinstitute.org). Upon verification you will receive a confirmation of receipt.

## FDP Examination

The FDP examination, administered twice annually, is a four-hour computer-administered examination that is offered at test centers throughout the world. The FDP examination is comprised of 75 multiple choice questions weighted as 65% of the total points and two to three constructed response questions (multi-part essay type) weighted as 35% of the total points. The FDP exam will not contain any Python or R programming questions.

The FDP examination is based on this study guide, which is organized to facilitate quick learning and easy retention. Each topic is structured around learning objectives and keywords that define the content to be tested on the exam. The learning objectives and keywords are an important way for candidates to organize their study, as they form the basis for examination questions. All learning objectives reflect the content in the FDP curriculum, and all examination questions are written to address the learning objectives directly.

For additional information about the FDP examination, please see the Candidate's Handbook, which can be found on the FDP Institute website.

## The FDP Curriculum: Outline

Candidates for the FDP Charter will have to enroll in the self-study program created by the FDP Institute and follow its carefully designed study guide. To become an FDP Charterholder, candidates must pass the FDP exam and submit their certificates of learning of the required online classes. The rest of this document discusses the FDP curriculum. Below is the outline of the curriculum:

Topics	Approximate Weight %
1. Introduction to Data Science & Big Data	5-10
2. Machine Learning: Introduction to Algorithms	5-10
3. Machine Learning: Regression, Support Vector Machine & Time Series Models	5-10
4. Machine Learning: Regularization, Regression Trees, Random Forest & Overfitting	5-10
5. Machine Learning: Classification & Clustering	5-10
6. Machine Learning: Performance Evaluation, Backtesting & False Discoveries	5-10
7. Data Mining & Machine Learning: Naïve Bayes & Text Mining	5-10
8. Big Data & Machine Learning: Ethical & Privacy Issues	5-10
9. Big Data & Machine Learning in the Financial Industry	30-50

## Other Study Tools and Resources

In addition to this study guide and candidate's handbook, the FDP Institute website directs you to the readings that are covered in the curriculum. The readings are detailed below by topic area and include textbooks, often used across topics, as well as several individual articles that are usually topic-specific. Both types of readings can be purchased from Amazon or the publisher, and whenever possible, they are posted on the FDP Institute website. They will be freely available to registered candidates.

### **Page Number References for Keywords**

For candidate's convenience, a set of six articles published by PMR Journals is provided in one collection titled Big Data & Machine Learning in the Financial Industry: Readings for the Financial Data Professional Exam and is available at a discounted price of \$99 for registered candidates. In this collection, there are two sets of page numbers: one corresponding to the collection's table of contents, and one corresponding to each article's page number in the original journal. The page numbers appearing next to the keywords refer to the page numbers as they appeared in the original article.

### **Sample Questions**

Candidates registered the March/April 2021 FDP exam have unlimited access to the sample questions to check their level of understanding of the course material.

## The FDP Curriculum: Reading List

### Topic 1: Introduction to Data Science & Big Data

- Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc. Chapters 1 & 2. Retrieved from <https://www.oreilly.com/library/view/data-science-for/9781449374273/> and <https://www.amazon.com/Data-Science-Business-Data-Analytic-Thinking-ebook/dp/B00E6EQ3X4>  
Make sure that you review the errata page of the book: <https://www.oreilly.com/catalog/errata.csp?isbn=0636920028918>
- Dannemiller, D. and R. Kataria. (2017). Alternative data for investment decisions: Today's innovation could be tomorrow's requirement by Deloitte Center for Financial Services. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-fsi-dcfs-alternative-data-for-investment-decisions.pdf>
- Gajjaria, A. (2018). Collective intelligence investing: Alpha generation via alternative data brings new risks by Deloitte Center for Financial Services. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-dcfs-im-collective-intel-investing.pdf>
- Quandl (2017). Sell Your Data to Wall Street: The comprehensive guide to monetizing your data assets for professional investors. Retrieved from <https://www.quandl.com/monetize-alternative-data/guide-to-selling-data-to-wall-street.pdf>

### Topic 2: Machine Learning: Introduction to Algorithms

- James, G., D. Witten, T. Hastie and R. Tibshirani. (2013). An Introduction to Statistical Learning: with applications in R. New York, NY: Springer. Chapters 1, 2.1 & 2.2  
<http://www-bcf.usc.edu/~gareth/ISL/> <https://www.amazon.com/Introduction-Statistical-Learning-Applications-Statistics/dp/1461471370>. Make sure that you review the errata page of the book: <http://faculty.marshall.usc.edu/gareth-james/ISL/errata.html>
- Nielsen, M. A. (2015). Using Neural Networks to Recognize Handwritten Digits. In Neural Networks and Deep Learning, Determination Press.  
Retrieved from <http://neuralnetworksanddeeplearning.com/chap1.html>

### Topic 3: Machine Learning: Regression, Support Vector Machine & Time Series Models

- Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc., Chapters 3 & 4. For links to the book and errata pages, see *Topic 1*
- James, G., D. Witten, T. Hastie and R. Tibshirani. (2013). An Introduction to Statistical Learning: with applications in R. New York, NY: Springer. Chapter 3, Sections 1-3. For links to the book and errata pages, see *Topic 1*
- Aas, K. and X. K. Dimakos. (2004). Statistical modeling of financial time series: An introduction. Oslo Norway: Norwegian Computing Center. Retrieved from <https://www.nr.no/files/samba/bff/SAMBA0804.pdf> **Sections 1-4**

## The FDP Curriculum: Reading List

### Topic 4: Machine Learning: Regularization, Regression Trees, Random Forest & Overfitting

- Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc., Chapter 5. For links to the book and errata pages, see *Topic 1*
- James, G., D. Witten, T. Hastie and R. Tibshirani. (2013). An Introduction to Statistical Learning: with applications in R. New York, NY: Springer. Chapters 6 & 8, Sections 6.1, 6.2, 8.2 & 8.2. For links to the book and errata pages, see *Topic 1*

### Topic 5: Machine Learning: Classification & Clustering

- Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc., Chapter 6 & 7. For links to the book and errata pages, see *Topic 1*

### Topic 6: Machine Learning: Performance Evaluation, Backtesting & False Discoveries

- Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc., Chapter 8. For links to the book and errata pages, see *Topic 1*
- Arnott, R., C. B. Harvey, and H. Markowitz. (2019). A Backtesting Protocol in the Era of Machine Learning. Journal of Financial Data Science, 1(1), 64-74.  
DOI: <https://doi.org/10.3905/jfds.2019.1.064>
- Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science, London, U.K.: Royal Society Open Science, Retrieved from <https://royalsocietypublishing.org/doi/full/10.1098/rsos.140216>
- López de Prado, M. (2019). A Data Science Solution to the Multiple-Testing Crisis in Financial Research. Journal of Financial Data Science, 1(1), 99-110.  
DOI: <https://doi.org/10.3905/jfds.2019.1.099>

### Topic 7: Data Mining & Machine Learning: Naïve Bayes & Text Mining

- Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc., Chapters 9 & 10. For links to the book and errata pages, see *Topic 1*
- Jurafsky, D. and J. Martin. (2018). Chapter 4. Naïve Bayes and Sentiment Classification, In Speech and Language Processing.  
Retrieved from <https://web.stanford.edu/~jurafsky/slp3/4.pdf>

## The FDP Curriculum: Reading List

### Topic 8: Big Data & Machine Learning: Ethical & Privacy Issues

- Institute of Business Ethics. (2016, June). Business Ethics and Big Data (IBE Issue 52). London, U.K.
- Institute of Business Ethics. (2018, January). Business Ethics and Artificial Intelligence (IBE Issue 58). London, U.K.
- Institute of Business Ethics. (2018, May). Beyond Law: Ethical Culture and GDPR (IBE Issue 62). London, U.K.
- Loukides, M., M., H. Mason and DJ. Patil. Ethics and Data Science  
<https://www.amazon.com/Ethics-Data-Science-Mike-Loukides-ebook/dp/B07GTC8ZN7>  
<https://www.oreilly.com/library/view/ethics-and-data/9781492043898/>  
<https://www.ebooks.com/en-us/book/209925972/Ethics-and-Data-Science/-Mike-Loukides>

### Topic 9: Big Data & Machine Learning in the Financial Industry

- Deloitte. (2018) Artificial intelligence. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/deloitte-analytics/deloitte-nl-data-analytics-artificial-intelligence-whitepaper-eng.pdf>
- AQR Portfolio Group Solutions (2019). Can machines “learn” finance? Retrieved from <https://www.aqr.com/Insights/Research/Journal-Article/Can-Machines-Learn-Finance>
- Deloitte. (2019) Artificial intelligence: The next frontier for investment management firms. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/fsi-artificial-intelligence-investment-mgmt.pdf>
- Financial Stability Board. (2017) Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications. Retrieved from <http://www.fsb.org/wp-content/uploads/PO11117.pdf>
- Monk, A., M. Prins, and D. Rook. (2019). Rethinking Alternative Data in Institutional Investment. Journal of Financial Data Science, 1(1), 14-31.  
DOI: <https://doi.org/10.3905/jfds.2019.1.1.014>
- Simonian, J., C. Wu, D. Itano and V. Narayanan. (2019). A Machine Learning Approach to Risk Factors: A Case Study Using the Fama–French–Carhart Model. Journal of Financial Data Science, 1(1), 32-44. DOI: <https://doi.org/10.3905/jfds.2019.1.032>
- Rasekhschaffe, K. and R. Jones. (2019). Machine Learning for Stock Selection. Financial Analyst Journal, 13 June 2019 Volume 75, Issue 3.  
Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3330946](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3330946)
- Gu, S., B. Kelly, and D. Xiu. (2018). Empirical Asset Pricing via Machine Learning. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3281018](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281018)  
<https://www.bryankellyacademic.org/> <http://dachxiu.chicagobooth.edu/download/ML.pdf>.  
**The paper posted on the FDP Institute’s website will be the version used for exam questions.**

**Topic 9: Big Data & Machine Learning in the Financial Industry *continued***

- López de Prado, M. (2018). The 10 Reasons Most Machine Learning Funds Fail. The Journal of Portfolio Management, 44 (6) 120-133; DOI: <https://doi.org/10.3905/jpm.2018.44.6.120>
- Harvey, C. R. and Y. Liu. (2014). Evaluating Trading Strategies. [Special 40th Anniversary Issue]. The Journal of Portfolio Management, 40(5), 108-118.  
DOI: <https://doi.org/10.3905/jpm.2014.40.5.108>
- Zappa, D., M. Borrelli, G.P. Clemente, N. Savelli. Text Mining In Insurance: From Unstructured Data To Meaning [https://www.variancejournal.org/articlespress/articles/Text\\_Mining-Zappa-Borrelli-Clemente-Savelli.pdf](https://www.variancejournal.org/articlespress/articles/Text_Mining-Zappa-Borrelli-Clemente-Savelli.pdf)

## Example of Learning Objectives

### Topic 1. Introduction to Data Science & Big Data

#### Readings

- 1.1 Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc. Chapters 1 & 2. Retrieved from <https://www.oreilly.com/library/view/data-science-for/9781449374273/> and <https://www.amazon.com/Data-Science-Business-Data-Analytic-Thinking-ebook/dp/B00E6EQ3X4>.  
*Make sure that you review the errata page of the book:*  
<https://www.oreilly.com/catalog/errata.csp?isbn=0636920028918>
- 1.2 Dannemiller, D. and R. Kataria. (2017). Alternative data for investment decisions: Today's innovation could be tomorrow's requirement by Deloitte Center for Financial Services Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-fsi-dcfs-alternative-data-for-investment-decisions.pdf>
- 1.3 Gajjaria, A. (2018). Collective intelligence investing: Alpha generation via alternative data brings new risks by Deloitte Center for Financial Services. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-dcfs-im-collective-intel-investing.pdf>
- 1.4 Quandl (2017). Sell Your Data to Wall Street: The comprehensive guide to monetizing your data assets for professional investors. Retrieved from <https://www.quandl.com/monetize-alternative-data/guide-to-selling-data-to-wall-street.pdf>

#### Reading 1.1 Provost, F. and T. Fawcett. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc. Chapters 1 & 2.

##### Keywords

*Data mining (p. 2)*

*Data science (p. 4)*

*Churn (p. 4)*

*Data-driven decision making (p. 5)*

*Data engineering (p. 5, 7)*

*Data-analytic thinking (p. 12)*

*Target (p. 24)*

*Label (p. 24)*

*Unsupervised data mining (p. 24)*

*Supervised data mining (p. 25)*

#### Learning Objectives

Demonstrate proficiency in the following areas:

##### 1.1.1 Data analytic thinking (Ch. 1)

*For example:*

- A. Discuss the ubiquity of data opportunities.
- B. Compare and contrast data science, engineering, and data-driven decision making.
- C. Explain data and data science capability as a strategic asset.

## Example of Learning Objectives

### 1.1.2 Business problems and data science solutions (Ch. 2)

*For example:*

- A. Describe how one transitions from business problems to data mining tasks.
- B. Compare supervised methods to unsupervised methods.
- C. Describe the difference between data mining and using the results of data mining.
- D. Describe key aspects of the data mining process, including business understanding, data understanding, data preparation, modeling, and evaluation.

**Reading 1.2 Dannemiller, D. and R. Kataria. (2017). Alternative data for investment decisions: Today's innovation could be tomorrow's requirement by Deloitte Center for Financial**

#### Keywords

*Diffusion of innovations curve (p. 2)*

*Innovators (p.3)*

*Early adopters (p.3)*

*Early majority (p.3)*

*Late majority (p.3)*

*Laggards (p.3)*

*Big data analytics (BDA) (p. 10)*

*Cognitive technologies (p. 12)*

### Learning Objectives

Demonstrate proficiency in the areas of:

#### 1.2.1 Adoption of alternative data

*For example:*

- A. Discuss investment managers in terms of their places on the diffusion of innovations curve, as discussed in Alternative data for investment decisions: Today's innovation could be tomorrow's requirement.

#### 1.2.2 The risks and rewards of alternative data for investment decisions

*For example:*

- A. Identify components of data risk.
- B. Identify four potential risks that may relate to incorporating alternative data in investment selection processes.
- C. Identify three risks that late adopters of alternative data in the investment management process face.

#### 1.2.3 Advanced technologies for alpha

*For example:*

- A. Identify two major types of technologies that investment management firms have started to use to realize alternative data's potential.

## Learning Objectives *continued*

- B. Describe the big data analytics (BDA) groups that are most aligned to support alpha generation.
- C. Describe the categories of cognitive technologies that are most aligned to support alpha generation.

### 1.2.4 Alternative data vendor profiles

*For example:*

- A. Identify the focus of the alternative data vendor profiled.
- B. Identify the four fundamental questions the alternative data vendor profiled asks to control the quality of alternative data.

### 1.2.5 The future of alternative data

*For example:*

- A. Identify the segment of alternative data adoption by investment managers (innovators, early adopters, early majority, late majority, laggards) expected within the next five years.

## Reading 1.3 Gajjaria, A. (2018). Collective intelligence investing: Alpha generation via alternative data brings new risks by Deloitte Center for Financial Services

### Keywords

<i>Collective Intelligence Investing (p. 1)</i>	<i>Harvest Exchange (p. 3)</i>
<i>Open communities (p. 3)</i>	<i>SumZero (p. 3)</i>
<i>Crowdsourcing platforms (p. 3)</i>	<i>Estimize (p. 3)</i>
<i>Digital expert contribution networks (p. 3)</i>	<i>Quantopian (p. 3)</i>
<i>Digital expert communication networks (p. 3)</i>	<i>Material nonpublic information risk (p. 5)</i>
<i>Seeking Alpha (p. 3)</i>	<i>Model risk (p. 5)</i>
<i>eToro (p. 3)</i>	<i>Information security risk (p. 5)</i>
<i>StockTwits (p. 3)</i>	<i>Data integrity risk (p. 6)</i>

## Learning Objectives

Demonstrate proficiency in the areas of:

### 1.3.1 Warming up to collective intelligence investing (CII)

*For example:*

- A. Describe the term collective intelligence investing.

### 1.3.2 Getting started with alternative data and CII

*For example:*

- A. Identify considerations and corresponding actions for traditional investment managers adopting alternative data and collective intelligence investing.

## Learning Objectives *continued*

### 1.3.3 Balancing the risks and rewards of CII from different platform types

*For example:*

- A. Describe four key types of collective intelligence communities: open communities, digital expert contribution networks, digital expert communication networks, and crowdsourcing platforms.
- B. Identify investment-related information that can be found in a specified type of CII community.
- C. Identify examples of a specified type of CII community, e.g., Seeking Alpha is an open community.
- D. Describe five key risks associated with CII.
- E. Describe a CII community in terms of the level of information structure, membership diversity, and risk exposure to each of the five key risks associated with CII.

### 1.3.4 The road ahead for investment managers

*For example:*

- A. Describe three steps investment manager firms could adopt before jumping wholeheartedly into CII for a potentially smoother takeoff.

### Reading 1.4 Quandt (2017). **Sell Your Data to Wall Street: The comprehensive guide to monetizing your data assets for professional investors**

**Keywords**

*Alpha (p. 2)*

*Buy side (p. 4)*

*Quantitative hedge funds (p. 4)*

*Contemporaneous correlation (p. 7)*

*Leading indicator (p. 7)*

*Table stakes (p. 9)*

*Data hunters (p. 12)*

*Personally identifiable information (PII) (p. 14)*

### Learning Objectives

Demonstrate proficiency in the areas of:

#### 1.4.1 Understanding why “Wall Street” wants data

*For example:*

- A. Define the buy-side.
- B. Identify the best customer for alternative data on the buy-side.
- C. Describe what alternative data provides the best customer on the buy-side.

#### 1.4.2 Building a data product

*For example:*

- A. Describe four stages of the production process.

## Learning Objectives *continued*

- B. Assess data quality and describe the criteria used.
- C. Evaluate the predictive power of a data set and describe the criteria used.
- D. Determine the commercial value of a data set and describe factors considered.
- E. Describe factors that go into packaging data for consumption.

### 1.4.3. Delivering your data

*For example:*

- A. Describe data delivery capabilities and features required by Wall Street customers.

### 1.4.4 Marketing and selling your data

*For example:*

- A. Identify sales and marketing strategies.
- B. Describe good data collateral.
- C. Describe the sales process.

### 1.4.5 Protecting yourself and your data

*For example:*

- A. Describe the concept of personally identifiable information (PII) and Wall Street's view of it.
- B. Describe the ability to sell data anonymously.

### 1.4.6 Pricing your data

*For example:*

- A. Describe what indicators and data attributes are related to the value of the data.
- B. Calculate the value of a data set given the assets under management (AUM) of a hedge fund and the expected excess returns the dataset is expected to generate.
- C. Describe strategies to balance the trade-offs involved in selling high-priced data to few clients versus more affordable data to many clients.

### 1.4.7 Case studies

*For example:*

- A. Illustrate how alternative data can help managers produce alpha when trading auto stocks.
- B. Illustrate how alternative data can help produce alpha when trading commodities.
- C. Illustrate how alternative data can help produce alpha when trading stocks.

**Create your free FDP Profile and gain access to the complete FDP Candidate Study Guide**

## Action Words

In each of the above learning objectives, action words are used to direct your study focus. Below is a list of all action words used in this study guide, along with definitions and two examples of usage, in a question example and a description. Should you not understand what is required for any learning objective, we suggest you refer to the table below for clarification.

**NOTE: The question examples in this table are NOT sample questions for the current exam.**

Term	Definition	Question Example	Example of Term Use
<b>Analyze</b>	Study the interrelations	George has identified an opportunity for a convertible arbitrage reverse hedge. What risks are associated with this hedge?  A. The convertible may remain overvalued, causing the positive cash flow to harm the position's return profile.  B. The short convertible may be called in and the position must be delivered, forcing the hedge to be unwound at an inopportune time.  C. The implied volatility may decrease, lowering the bond's value.	You have to <b>analyze</b> the positions and factors impacting them.  <b>Correct Answer: B</b>
<b>Apply</b>	Make use of  <b>Note:</b> If you are asked to apply a model to data, you will be expected to have the appropriate equation memorized, unless the question also contains the action word "recognize"..	Alicia Weeks, CFA, Real Estate Investment Advisor, works in an Asian country where there are no securities laws or regulations. According to CFA Institute Standard I, Fundamental Responsibilities, Alicia:  A. Must adhere to the standards as defined in a neighboring country that has the strictest laws and regulations.  B. Need not concern herself with ethics codes and standards.  C. Must adhere to the CFA Institute's codes and standards.	You have to <b>apply</b> CFA Institute Standard I to find the correct answer.  <b>Correct Answer: C</b>
<b>Argue</b>	Prove by reason or by presenting the associated pros and cons; debate	Why did the shape of the supply curve for venture capital funds change after 1979?	You have to describe how the curve has changed AND <b>argue</b> why it changed by providing reasons and supporting the reasons with statements of facts (e.g., change in regulations).
<b>Assess</b>	Determine importance, size, or value	How are lower capital gains taxes expected to impact firm commitments?  A. Through increased supply of capital, firm commitments are expected to rise.  B. Through decreased supply of capital, firm commitments are expected to rise.  C. Through decreased after-tax return on venture investments, firm commitments are expected to rise.	You must <b>assess</b> the significance of the change in the tax rate for firm commitments.  <b>Correct Answer: A</b>

## Action Words

Term	Definition	Question Example	Example of Term Use
<b>Calculate</b>	Determine a value mathematically  <b>Note:</b> You will be expected to have the appropriate equation memorized, unless the question also contains the action word "recognize".	Consider a set of 100 people. Eighty percent have feature A and twenty percent do not have feature A. What is the entropy for this set?  A. 0.72 B. 0.88 C. 0.93	You have to <b>calculate</b> entropy based on the given probabilities.  <b>Correct Answer: A</b>
<b>Compare</b>	Describe similarities and differences	Which of the following least accurately compares the Sharpe and Treynor ratios?  A. Both ratios contain excess return in the numerator.  B. Both ratios express a measure of return per unit of some measure of risk.  C. The Sharpe ratio is the inverse of the Treynor ratio	You must <b>compare</b> the ratios based on their most important similarities and their most important differences.  <b>Correct Answer: C</b>
<b>Compare and Contrast</b>	Examine in order to note similarities or differences	A comparison of monthly payments and loan balances of a constant payment mortgage with a constant amortization mortgage with the same loan terms will show that:  A. The initial payment will be the same.  B. The payments of the constant payment mortgage are initially greater than those of the constant amortization mortgage, but at some point the payments of the constant payment mortgage become less.  C. The present value of the payment streams of the two loan types are the same.	You must <b>compare</b> indices to arrive at the answer.  <b>Correct Answer: C</b>
<b>Construct</b>	Make or form by combining or arranging parts or elements	A reverse convertible arbitrage hedge consists of a:  A. Short convertible position plus a long position in the stock. B. Short convertible position plus a put option on the stock. C. Long convertible position plus a put option on the stock.	You must <b>combine</b> positions to construct the hedge.  <b>Correct Answer: A</b>
<b>Contrast</b>	Expound on the differences	Which of the following best characterizes a difference between value at risk (VaR) and modified VaR?  A. Modified VaR is expressed as a percent while VaR is a dollar value.  B. Modified VaR uses a user defined confidence interval while VaR uses a 99% interval.  C. Modified VaR incorporates non-normality while traditional VaR assumes normality	You have to <b>contrast</b> the assumptions of the first model to those of the second model so that the differences are clear.  <b>Correct Answer: C</b>

## Action Words

Term	Definition	Question Example	Example of Term Use
<b>Define</b>	State the precise meaning	The interest rate charged by banks with excess reserves at a Federal Reserve Bank to banks needing overnight loans to meet reserve requirements is called the:  A. Prime rate.  B. Discount rate.  C. Federal funds rate.	You must <b>define</b> , in this case, the federal funds rate.  <b>Correct Answer: C</b>
<b>Describe</b>	Convey or characterize an idea	Which of the following words best describes expected return?  A. Spread  B. Average  C. Spread squared	You need to choose the word that best <b>describes</b> the concept from a list.  <b>Correct Answer: B</b>
<b>Differentiate</b>	Constitute the distinction between; distinguish	What type of convertible hedge entails shorting a convertible and going long in the underlying stock?  A. Reverse hedge  B. Call-option hedge  C. Traditional convergence hedge	You must <b>differentiate</b> one type of hedge from another.  <b>Correct Answer: A</b>
<b>Discuss</b>	Examine or consider a subject	Discuss the limitations of private equity data.	You must present a <b>discussion</b> of a set of ideas in a list or paragraph.
<b>Explain</b>	Illustrate the meaning	1. Explain why return on assets (ROA) rather than return on equity (ROE) might be the preferred measure of performance in the case of hedge funds.  or  2. Which of the following best explains risk from the standpoint of investment?  A. Investors will lose money.  B. Terminal wealth will be less than initial wealth.  C. More than one outcome is possible.	1. You must place a series of thoughts together as an <b>explanation</b> of a term or issue.  2. You need to identify the term that best <b>explains</b> a term or issue.  <b>Correct Answer: C</b>
<b>Identify</b>	Establish the identity	The investments that have historically performed best during periods of recession are:  A. Commodities.  B. Treasury bills.  C. Stocks and bonds.	You must <b>identify</b> the term that best meets the criterion of the question  <b>Correct Answer: C</b>
<b>Illustrate</b>	Clarify through examples or comparisons	For two types of convergence hedges, what situations present profitable opportunities, how are the hedges set up, and what are the associated risks?	You must provide an example for each hedge or compare the two to <b>illustrate</b> how they work.

## Action Words

Term	Definition	Question Example	Example of Term Use
<b>Identify</b>	Establish the identity	The investments that have historically performed best during periods of recession are:  A. Commodities.  B. Treasury bills.  C. Stocks and bonds.	You must <b>identify</b> the term that best meets the criterion of the question          <b>Correct Answer: C</b>
<b>Illustrate</b>	Clarify through examples or comparisons	For two types of convergence hedges, what situations present profitable opportunities, how are the hedges set up, and what are the associated risks?	You must provide an example for each hedge or compare the two to <b>illustrate</b> how they work.
<b>Interpret</b>	Explain the meaning	Your certificate of deposit will mature in one week, and you are considering how to invest the proceeds. If you invest in a 30-day CD, the bank will pay you 4% interest. If you invest in a 2-year CD, the bank will pay you 6% interest.  You should choose the:  A. 2-year CD if you expect that interest rates will fall in the future  B. 30-day CD, no matter what you expect interest rates to do in the future.  C. 2-year CD, no matter what you expect interest rates to do in the future.	You must <b>interpret</b> the features of an investment scenario.          <b>Correct Answer: A</b>
<b>List</b>	Create a series of items	List the determinants of real interest rates.	You must differentiate from a <b>list</b> those items that are consistent with the question.
<b>Outline</b>	Summarize tersely	Which of the following best characterizes the steps in computing a geometric mean return based on a series of periodic returns from $T$ time periods?  A. Add one to each return, add them together, divide by $T$ and subtract one.  B. Add one to each return, multiply them together, take the $T$ th root and subtract one.  C. Add one to each return, multiply them together, divide by $T$ and subtract one.	You must <b>outline</b> the study's most important findings rather than explain them in detail.          <b>Correct Answer: B</b>

## Action Words

Term	Definition	Question Example	Example of Term Use
<p><b>Recognize</b></p>	<p>Recall the purpose of a given equation or term, and its name when appropriate.</p> <p><b>Note:</b> When the action word “recognize” is used and applied to an equation, the equation will be provided within the question stem or the correct answer choice.</p>	<p>What is the following equation called and used for in the context of artificial neural networks?</p> $\sigma(z) \equiv \frac{1}{1 + e^{-z}}$ <p>A. It is called a neuron and used to make a NAND gate.</p> <p>B. It is called a sigmoid function and is used to model sigmoid neurons that better enable learning than perceptrons do.</p> <p>C. It is called a perceptron which is used to create a smoother function than a logistic function.</p>	<p>You must <b>recognize</b> that this is a sigmoid function, also referred to as a logistic function, which is used as the basis for a sigmoid neuron.</p> <p><b>Correct Answer: B.</b></p>
<p><b>Relate</b></p>	<p>Show or establish logical or causal connection</p>	<p>Which of the following effects does NOT help to explain growth in the venture capital industry?</p> <p>A. Amendments to the prudent man rule</p> <p>B. The rise of limited partnerships as an organizational form</p> <p>C. Decline in the valuations of small capitalization stocks</p>	<p>You must <b>relate</b> effects or factors (e.g., the prudent man rule) to another result or concept (e.g.) growth in an industry.)</p> <p><b>Correct Answer: C</b></p>