

Analytics mindset**BioPhirma**

The following files are used for this case and are presented in the order of use by the type of file.



Case study and solutions

- Analytics_mindset_case_studies_BioPhirma.docx**
- Analytics_mindset_case_studies_BioPhirma.docx**
- Analytics_mindset_case_study_solutions_BioPhirma.docx**
- Analytics_mindset_case_study_solutions_BioPhirma.pdf**



Data sets

- Analytics_mindset_case_studies_BioPhirma.xlsx**



Analytics workbooks

Alteryx packaged workbooks:

- Part 2: ETL
 - Analytics_mindset_case_study_solutions_BioPhirma_Part2.yxzp**
- Part 3a: Employee analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3a.yxzp**
- Part 3b: Vendor analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3b.yxzp**
- Part 3c: Month and department analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3c.yxzp**
- Part 3d: Transaction amount violation analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3d.yxzp**
- Part 3e: Split transaction violation analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3e.yxzp**
- Part 3f: Unapproved vendor analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3f.yxzp**
- Part 3g: Sign-off analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3g.yxzp**
- Part 3h: Day of week analysis
 - Analytics_mindset_case_study_solutions_BioPhirma_Part3h.yxzp**

Part 3i: Key word analysis

[Analytics_mindset_case_study_solutions_BioPhirma_Part3i.yxzp](#)

Part 3j: TF-IDF analysis

[Analytics_mindset_case_study_solutions_BioPhirma_Part3j.yxzp](#)



Videos

How-to videos: **Note that video links are removed for the external distribution of this EYARC Access newsletter. See the user guide on the EYARC site to access the video links.**

Part 2: ETL

[Analytics mindset case study solutions BioPhirma Part2](#)

Part 3: Apply appropriate data analytics techniques

[Analytics mindset case study solutions BioPhirma Part3a](#)

[Analytics mindset case study solutions BioPhirma Part3b](#)

[Analytics mindset case study solutions BioPhirma Part3c](#)

[Analytics mindset case study solutions BioPhirma Part3d](#)

[Analytics mindset case study solutions BioPhirma Part3e](#)

[Analytics mindset case study solutions BioPhirma Part3f](#)

[Analytics mindset case study solutions BioPhirma Part3g](#)

[Analytics mindset case study solutions BioPhirma Part3h](#)

[Analytics mindset case study solutions BioPhirma Part3i](#)

[Analytics mindset case study solutions BioPhirma Part3j](#)

Overview

This case is designed for an accounting information systems, auditing, fraud or data analytics course. The case is designed to be flexible so that it can be implemented in class, used as homework, or as a student project or part of an exam.

In this case, students are asked to assume the role of an advisor to BioPhirma, a global corporation based in Atlanta, Georgia. BioPhirma is an innovative health care leader that conducts extensive research and development activities to invent and test medications and treatments for a wide array of human and animal medical conditions. Students are asked to advise BioPhirma on compliance with its purchasing card (P-card) policies and to help investigate a whistle-blower tip about possible fraud being committed through P-card transactions.

A key feature of this case is that it walks students through the entire analytics mindset approach. The students are able to practice developing their analytics mindset with rich, real-world data. As a reminder, an analytics mindset is the ability to:

- ▶ Ask the right questions
- ▶ Extract, transform and load relevant data (i.e., the ETL process)

Analytics mindset case studies – BioPhirma – User guide

© 2020 Ernst & Young Foundation (US). All Rights Reserved.

SCORE no. 08247-201US_3

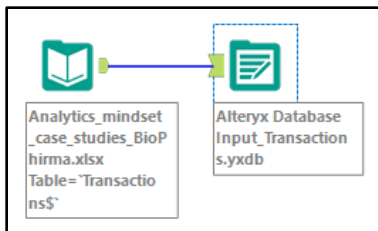
- ▶ Apply appropriate data analytics techniques
- ▶ Interpret and share results with stakeholders

This is a four-part case. The following is a brief overview of each part.

Part 1: In this part, students are given background about BioPhirma and its P-card transaction process. Students are also given insights into the current use of data and analytics in managing fraud risk. An appendix is provided that provides BioPhirma’s P-card policies and also an overview of a specific text-mining analytics technique, Term Frequency-Inverse Document Frequency (TF-IDF). Students are asked to think about the right questions they need to ask. This includes a focus on identifying risks related to the P-card transactions, controls that are (or are not) in place and the types of data analyses that could be considered to test these controls. Students are also asked to make a recommendation to BioPhirma’s management about additional controls or process changes that could be considered to address the risk related to its P-card transactions, even prior to performing any data analysis.

Part 2: This part focuses on the ETL process. The data extraction process already has been completed for the students. Students will receive one Excel data file with three tabs (Transactions, Approved Vendors List and Authorized P-card Users) that includes all of the data needed. Students will be provided with the details of each data field within the data set in the case background. Students are asked to complete select data validations to make certain that their data file is complete. Students are required to load their data correctly into the recommended tool, Alteryx. This loading, and any additional data transformation needed for each analysis, will be shown in Part 3.

Note: As the data file was not exceptionally large, we did not include a requirement for students to convert each tab in the Excel data file into a separate Alteryx database file (.yxmd file type) to use as the input file for each workflow. Doing so typically provides more efficiency when a workflow is run. You can require students to perform this step by creating a two-step workflow with the Input Data tool and selecting the Excel file and then using the Output Data tool with a file type of Alteryx Database, as shown below.



Part 3: Students are asked to perform a series of analyses using Alteryx to provide insights into the compliance with BioPhirma’s P-card policies and to identify any potential fraudulent transactions. Note that you can ask students to perform all of these analyses, or selectively based on your learning objectives. You can also require them to create an additional analysis of their own as this set of analyses is not comprehensive in addressing all possible risks and insights needed. In terms of complexity, Parts 3i and 3j are the most complex. There is a chart following that indicates the tools used in each Alteryx workflow to provide you with an overview. Please be aware that there could be many approaches to completing each workflow. The following analyses are included:

3a: Employee

3b: Vendor

3c: Month and department

3d: Transaction amount violation

3e: Split transaction violation

3f: Unapproved vendor

3g: Sign-off

3h: Day of week

3i: Key word: Note that you can discuss with students the key words provided and whether there are additional words they would like to add to the search and why these key words might be appropriate to target. Support for this might be based on research of typical key words that forensic accountants might seek to analyze based on the type of data source, or just based on their understanding of the company and industry.

3j: TF-IDF: Note that in the data cleansing for this analysis, Step 1.1, there is an option to remove unwanted characters with “Leading and Trailing Whitespace” selected. You might want to discuss with your students the results that would be generated by selecting “Punctuation,” as well. For example, if a P-card holder enters a word in the description, but it is entered with various punctuation following it, such as a semicolon or a period, the analysis would identify each of these separately due to the punctuation, rather than as the same word. Alternatively, punctuation might be very significant, such as a hyphen used in separating a series of numbers. This could be an example of when the students might initially perform their analysis and then further refine their data cleansing based on what they see. It is important to note that the selection of “Punctuation” in the data cleansing tool does not give an option to select specific punctuation characters; therefore, it is all or nothing. If students are interested in removing select punctuation, they could consider adding a formula to cleanse the data, such as REPLACE where a specific character is replaced with an empty string (“”).

Part 4: Students are asked to interpret the results and make appropriate recommendations. Students should be able to use their Alteryx workflows in a live setting to present their findings. Several of the analyses require students to use the Reporting tool in Alteryx to generate interactive visualizations. Additionally, any other findings can be browsed and data within can be sorted or filtered, as needed. Alternatively, you can require students to produce output files from the workflows, including images, Excel files and more.

Advanced preparation

It is recommended that you expose students to the definition and importance of the analytics mindset and related competencies prior to covering this case. The EYARC offers lecture notes, slides and a competency framework in the *Introduction to the analytics mindset* module that you can use for this purpose.

Data

The data for this case is real-world P-card data that includes individual purchase transactions (approximately 82,000), approved vendors (approximately 10,600) and P-card users (approximately 1,200). Select data has been cleansed for data privacy purposes.

Analytics tools

We have designed this case to be flexible with respect to the tools that can be used to perform the analyses. We provide the case and the solutions using Alteryx. This software is one of the leading self-service data analytics programs currently available. It is a powerful tool for the ETL process and data analytics, and is especially known for its spatial analytics tools. “Alteryx allows a single user to access various data sources, clean and prepare data, perform a variety of analyses and then deploy the results for consumption and to operationalize the insights discovered. It boasts visual workflows and an intuitive drag-and-drop interface that can eliminate the need to write code.”¹

The workflows allow the user to visually understand the steps that the data goes through for transformation or analysis. This workflow also provides a clear audit trail and creates a repeatable process. When using an Alteryx workflow, the initial data sources remain intact. Alteryx performs the transformation and generates a new output file, which helps maintain the integrity of the data, allows mistakes to be easily fixed and allows the same workflow to be used on multiple data sets (with the same fields and properties). Alteryx can process very large data sets very quickly. As the workflow is running, the user can see where any errors may be occurring and can monitor the progress as percentage-of-completion statistics are calculated throughout. As soon as the data is processed, a report is generated that shows if there are any errors in any fields (e.g., trailing white spaces) that may create problems for data analysis, as well as some basic visual analytics that show descriptive information about the processed data.

Access: A free trial of Alteryx and academic licensing is available at <https://www.alteryx.com/why-alteryx/alteryx-for-good>. Alteryx will provide individual licenses for students and faculty, as well as lab licenses for use in a computer lab or classroom. However, it is important to note that Alteryx only works on PCs at this time.

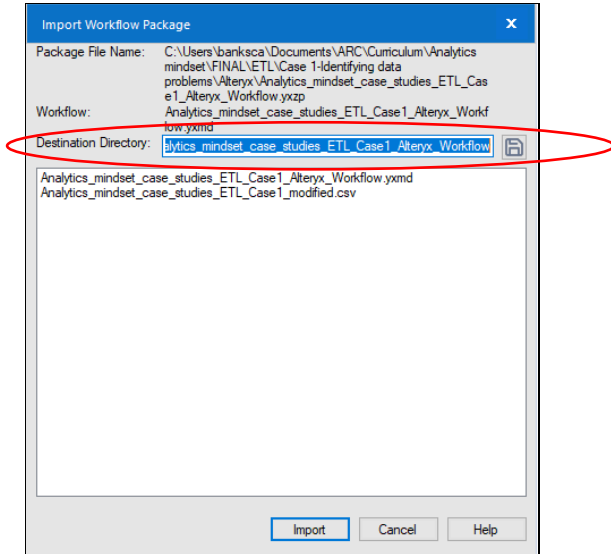
Community and training:

- ▶ Alteryx is intuitive and easy to learn. It boasts an active user community (<https://www.alteryx.com/community>) that openly shares workflow examples to help users identify potential solutions. Like other analytics tools, there are always many ways to organize data to achieve the desired solutions. By reviewing approaches of other data scientists, you and your students can develop more advanced skills and perspectives.
- ▶ Also within the Alteryx community, there are easy-to-follow tutorials as part of the Alteryx Academy (<https://community.alteryx.com/t5/Alteryx-Academy/ct-p/alteryx-academy>). You can learn most of the necessary ETL functions in a few hours using these tutorials.
- ▶ Further, there is a set of tutorials designed with the proficient Excel user in mind that walks you through basic Excel functions using Alteryx instead. As users become more advanced, they can participate in the Alteryx “weekly challenges” (<https://community.alteryx.com/t5/Weekly-Challenge/bd-p/weeklychallenge>), which provide difficult data analytics problems for the user community to solve.

Note that for Alteryx workflow solution files, these have been provided as a packaged workflow (.yxzp file type [Options > Export Workflow >]) that is a single, zipped file that includes the Alteryx workflow and all

¹ “Alteryx, Inc. Form 10-K,” *Securities and Exchange Commission website*, <https://www.sec.gov/Archives/edgar/data/1689923/000119312518073878/d530988d10k.htm>, accessed January 1, 2020.

of its dependencies (e.g., input file and naming convention for the output file). Students are asked to submit their completed packaged workflow as their deliverable using this file type as well, including their name in the file, to allow the workflow to be easily executed. When this packaged workflow is opened, you will be asked to provide a file directory where you want the files to be saved before you are able to run the workflow.



Annotation

One of the nice features of this tool is that it allows the user to document each step of the workflow through the use of annotations. The case does not include requirements for students to annotate each step of their workflow as written, although annotations have been provided in the solution set. You can add a requirement with a description of the level of detail that you would like to see, as desired.

How-to videos

How-to videos have been prepared to accompany this case, with links listed on page two. These links have *not* been included in the case solutions file. These videos can help instructors learn the Alteryx workflow steps and tools themselves and they can be provided directly to students to supplement in-class instruction and facilitate their completion of the case.

Alteryx workflow tool matrix

See the following page for an overview of the Alteryx tools used for each workflow in the case.



		Alteryx workflow															
Part	Analysis	In/out			Preparation						Join		Parse	Transform		Reporting	
2	ETL																
3a	Employee																
3b	Vendor																
3c	Month and department																
3d	Transaction amount violation																
3e	Split transaction violation																
3f	Unapproved vendor																
3g	Sign-off																
3h	Day of week																
3i	Key word																
3j	TF-IDF																