

## **Analytics mindset**

### **Fairview Inc.**

#### **Extract, transform and load (ETL)**

#### **Overview**

You work as a corporate accountant for Fairview Inc., a software-as-a-service (SaaS) company. Fairview Inc. uses a unique “high-human-touch” sales and servicing model. That is, Fairview Inc. sells only within certain cities and hires numerous sales agents and information technology (IT) agents in each city where it operates. The company then has these employees visit the customers and provide face-to-face sales and support.

The model has been successful so far and Fairview Inc. now operates in five cities in the Midwest of the United States: Indianapolis, Kansas City, Louisville, Nashville and St. Louis. Each city has a physical office staffed by a manager and customer support representatives. The office manager is responsible for and oversees all work done in the office. The customer support representatives field calls and send sales agents and IT service representatives to visit clients. When a customer support representative receives a call, they cycle through a list of employees in assigning workers to customers, assigning the employee who has not performed the task for the longest time. That way, every employee has an equal opportunity to visit customers. The customer service representative keeps a separate list of sales agents and IT service representatives.

Fairview Inc. pays employees for the time they spend visiting customers. The company also reimburses employees when they travel to a customer’s business since employees use their own vehicles for transportation. After the trip, the employee submits their mileage for reimbursement. The employee submits their home address, the customer address and their self-reported mileage that they traveled to visit the customer. The company then reimburses the employee using the Internal Revenue Service (IRS) standard mileage rate.

The CFO at Fairview Inc. wants to perform a comprehensive review of the reimbursement process. This includes having one team analyze the process to look for automation possibilities, having Internal Audit conduct an audit of the program and look for internal control deficiencies and potential fraud, and having another team analyze the process for efficiency gains. To perform the analyses, each group will need data. While each group will need a slightly different set and structure for the data, each will start with a basic data file. Your task is to prepare the basic data file for these analyses. As you prepare this file, you should make the process repeatable, so that when the IT Department pulls new files, your workflow will be able to automatically update and create a new basic data file.

The IT Department extracted the data from the company's IT systems and has prepared files that you will manipulate to prepare your final file. The IT Department has pulled together a list of the trips for the previous year. In total, employees sought reimbursements for 4,888 trips at a total cost of \$39,658.61. Note that the IT Department has pulled several versions of different files. Depending on the assignment noted below, you will use these different files. The files and an explanation are as follows:

- ▶ The file, **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx**, contains eight sheets of data. The sheets are as follows:
  - Trips: sheet lists the address of the employee, the customer address, the office of the employee and the time of the trip. The only information that is meaningful in the TimeOfTrip column is the actual time (sometimes programs add in date information to time fields, but this has no meaning in this data).
  - EmployeeData: sheet lists information about the employee, including their first and last name, employee address, office and their role (either Sales or IT).
  - CustomerData: sheet lists the name of the customer's company, the customer address and the office to which they belong.
  - Sheets: Indianapolis, KansasCity, Louisville, Nashville and StLouis: Sheets each contain a matrix that shows the customer and employee addresses and the distances between every address.
- ▶ **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL\_Indianapolis.csv** file. This file contains the same information as the Indianapolis sheet in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file. The only difference is that some columns of information may be duplicated.
- ▶ **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL\_KansasCity.csv** file. This file contains the same information as the Indianapolis sheet in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file. The only difference is that the employee address listed on the first row in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file is split onto two lines in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL\_KansasCity.csv** file.
- ▶ **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL\_StLouis.csv** file. This file contains the same information as the St. Louis sheet in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file. The only difference is that the employee addresses use delimiter and text qualifiers that are different from the rest of the file.

### **Analytics mindset**

Having and using an analytics mindset is critical in accounting and business. An analytics mindset is the ability to:

- ▶ Ask the right questions
- ▶ Extract, transform and load (ETL) relevant data
- ▶ Apply appropriate data analytics techniques

- ▶ Interpret and share the results with stakeholders

In this case, using an analytics mindset means properly analyzing the data files and deciding the correct way to perform the ETL procedures on the data.

## Required

This assignment is broken into three parts. The parts build upon each other — once you finish Part 1, you can edit that file to finish Part 2, which, in turn, can be edited to finish Part 3. Each of your solutions must be “automated.” In this way, if the company were to change any of the trip data, or were to change the mileage between employees and customers, your solution could be run again and to create the final file with minimal effort (i.e., manually editing the data in Microsoft Excel is not an acceptable solution).

### Part 1

You should only use the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file for this part. Transform the data so that your final output is an Excel file called “Solution Part 1\_StudentName.xlsx.” The first few rows of the file should look as follows:

	A	B	C	D	E	F	G	H	I	J	K
1	RowID	EmployeeID	EmpFullName	EmployeeAddress	EmpRole	CustomerID	CustomerName	CustomerAddress	Office	ValidatedMileage	TimeOfTrip
2	1	1130	Huzaifah Harrington	2166 Tower Grove Ave, St. Louis, MO, 63110	IT	101141	Bosco Inc	292 East Ave, Saint Louis, MO, 63119	St. Louis	6.8	10:35:00
3	2	1052	Hussain Mullins	8630 Haskell Ave, Kansas City, MO, 66109	Sales	100285	Jaskolski, Steuber and Herzog	6300 W 143rd St Ste 120, Overland Park, KS, 66223	Kansas City	25.1	12:45:00
4	3	1122	Susan Velasquez	1515 Partridge Ave, St. Louis, MO, 63133	Sales	101183	Blick Inc	11100 Olive Blvd, Saint Louis, MO, 63141	St. Louis	6.6	15:09:00
5	4	1118	Milana Sampson	2048 Castleman Dr, Nashville, TN, 37215	IT	100873	Lueilwitz-Bode	1606 16th Ave S, Nashville, TN, 37212	Nashville	3.7	16:05:00
6	5	1054	Jovan Yang	5755 Windsor Dr, Fairway, KS, 66205	Sales	100294	Kris-Muller	4741 Central St Ste 447, Kansas City, MO, 64112	Kansas City	2.8	14:06:00
7	6	1067	Mylee Boyer	401 N Sherwood Ave, Clarksville, IN, 47129	Sales	100488	Lebsack-Hartmann	4411 Peregrine Pl, Louisville, KY, 40241	Louisville	16.6	11:50:00
8	7	1130	Huzaifah Harrington	2166 Tower Grove Ave, St. Louis, MO, 63110	IT	101176	Dach-Marquardt	701 Convention Plz, Saint Louis, MO, 63101	St. Louis	5.7	11:57:00
9	8	1081	Elliott Reader	623 Keeton Ave, Old Hickory, TN, 37138	Sales	100676	Klocko, Romaguera and Fritsch	106 Mission Ct Ste 604, Franklin, TN, 37067	Nashville	31.2	17:59:00
10	9	1119	Viaan Leach	206 Meridian St, Nashville, TN, 37207	IT	101046	Lesch-Howell	1251 Lewisburg Pike, Franklin, TN, 37064	Nashville	25.8	16:58:00
11	10	1061	Kalem Nguyen	4501 NW Lakota Ct, Riverside, MO, 64150	Sales	100463	Ortiz-Huels	4520 Main St Fl 8th, Kansas City, MO, 64111	Kansas City	10.5	14:34:00

The columns are the same as in the sheets in the **Analytics\_mindset\_case\_studies\_FairviewInc.xlsx** file. The ValidatedMileage field (column J in the image above) is the distance listed on one of the city sheets between the employee and customer address.

- ▶ Save your solution file as “Solution Part 1\_StudentName” with whatever extension is used for the software you choose.

## Part 2

Part 2 builds on Part 1. Copy the solution to Part 1 and continue building on this copied file to build the solution to Part 2. Like Part 1, you should only use the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file for this part. Transform the data so that your final output is an Excel file called “Solution Part 2\_StudentName.xlsx.” The first few rows of the file should look as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	RowID	EmployeeID	EmpFullName	EmployeeAddress	EmpRole	CustomerID	CustomerName	CustomerAddress	Office	ValidatedMileage	TimeOfTrip	MinMileage	MaxMileage	MinEmpFullName	MaxEmpFullName
2	1	1130	Huzaifah Harrington	2166 Tower Grove Ave, St. Louis, MO, 63110	IT	101141	Bosco Inc	292 East Ave, Saint Louis, MO, 63119	St. Louis	6.8	10:35:00	1.4	17.3	Leigha Rodgers	Brax Harrison
3	2	1052	Hussain Mullins	8630 Haskell Ave, Kansas City, MO, 66109	Sales	100285	Jaskolski, Steuber and Herzog	6300 W 143rd St Ste 120, Overland Park, KS, 66223	Kansas City	25.1	12:45:00	5.5	27.1	Nina Dolan	Kalem Nguyen
4	3	1122	Susan Velasquez	1515 Partridge Ave, St. Louis, MO, 63133	Sales	101183	Blick Inc	11100 Olive Blvd, Saint Louis, MO, 63141	St. Louis	6.6	15:09:00	5.9	22.2	Jasmine Ferrell	Brax Harrison
5	4	1118	Milana Sampson	2048 Castleman Dr, Nashville, TN, 37215	IT	100873	Lueilwitz-Bode	1606 16th Ave S, Nashville, TN, 37212	Nashville	3.7	16:05:00	0.9	22	Loui Sharpe	Fern Sears
6	5	1054	Jovan Yang	5755 Windsor Dr, Fairway, KS, 66205	Sales	100294	Kris-Muller	4741 Central St Ste 447, Kansas City, MO, 64112	Kansas City	2.8	14:06:00	2.8	16.6	Brayden Krause	Hussain Mullins
7	6	1067	Mylee Boyer	401 N Sherwood Ave, Clarksville, IN, 47129	Sales	100488	Lebsack-Hartmann	4411 Peregrine Pl, Louisville, KY, 40241	Louisville	16.6	11:50:00	4.6	22.2	Bo Sanford	Ruben Bond
8	7	1130	Huzaifah Harrington	2166 Tower Grove Ave, St. Louis, MO, 63110	IT	101176	Dach-Marguardt	701 Convention Plz, Saint Louis, MO, 63101	St. Louis	5.7	11:57:00	3.3	19.4	Justine Ramirez	Cara Bonilla
9	8	1081	Eliott Reader	623 Keeton Ave, Old Hickory, TN, 37138	Sales	100676	Klocko, Romaguera and Fritsch	106 Mission Ct Ste 604, Franklin, TN, 37067	Nashville	31.2	17:59:00	2.2	35.5	Cody Macgregor	Rania Daniel
10	9	1119	Viaan Leach	206 Meridian St, Nashville, TN, 37207	IT	101046	Lesch-Howell	1251 Lewisburg Pike, Franklin, TN, 37064	Nashville	25.8	16:58:00	6.6	41.9	Cody Macgregor	Rania Daniel
11	10	1061	Kalem Nguyen	4501 NW Lakota Ct, Riverside, MO, 64150	Sales	100463	Ortiz-Huels	4520 Main St Fl 8th, Kansas City, MO, 64111	Kansas City	10.5	14:34:00	2.7	16.1	Brayden Krause	Hussain Mullins

The columns are the same as in the sheets in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file. The ValidatedMileage field (column J in the image above) is the distance listed on one of the city sheets between the employee and customer address. The fields MinMileage and MaxMileage contain the distance from the customer to the closest and farthest employee, respectively. The fields MinEmpFullName and MaxEmpFullName contain the name of the employee who is closest to or farthest from the customer, respectively. If two employees are the exact same distance away, you should always return the employee who is listed first alphabetically by the employee address (sorted in ascending order).

► Save your solution file as “Solution Part 2\_StudentName” with whatever extension is used for the software you choose.

## Part 3

Part 3 builds on Part 2 and, thus, Part 1. The output for Part 3 should look identical to the output for Part 2. In Part 3, the company realized that the CSV data files for Indianapolis, Kansas City and St. Louis are structured differently from what is listed in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file. Thus, your solution to Part 3 must use these three alternative CSV data files rather than the similar sheets in the **Analytics\_mindset\_case\_studies\_FairviewInc\_ETL.xlsx** file. The specific problems in each file are as follows:

► Indianapolis data:

- There are duplicate rows for 12 customers. A duplicate contains the exact same information for all information on the row.

▶ Kansas City data:

- The address of each employee is on the first two lines, instead of just the first line. On the first line is the street address, on the second line is the city, state and ZIP code information.

▶ St. Louis data:

- The data uses piped delimiters after the first row. However, the delimiters not only separate the mileage, but also the information in the customer address. That is, the street address, city, state and ZIP code are also separated by pipes.

Save your solution file as “Solution Part 3\_StudentName” with whatever extension is used for the software you choose.

**Overall deliverables to submit for grading:**

- ▶ The packaged analytics workbook solution for each part (i.e., three files). Include your name at the end of each of the file names.
- ▶ The Excel solution output file for each part (i.e., three files). Include your name at the end of each of the file names.