

Missing Money Matters - Fraud Analysis Case Study

By: Jarrett T. Bruno

Case Study Parameters

- WSDA Music is a fictional company that sells and distributes audio files ranging from music to audiobooks.
- Primary stakeholder: Adam Andrews, Manager of WSDA Music
- There is a discrepancy in the company's financials between the years 2011 and 2012

Deliverables

Utilize SQL to uncover the following:

- Get a list of suspects.
- Narrow the list
- Pinpoint the prime suspect(s)

About the data:

- Used SQLite for analysis
- Dataset contains 11 tables
- For our analysis we'll be using the 'customers', 'employees', and 'invoices' tables

Challenge 1

- How many transactions took place between 2011 and 2012?
- How much money did WSDA Music make between 2011 and 2012?

WSDA Music made \$1947.97 between 2011 and 2012.

```
-- 1. How many transactions took place between the years 2011 and 2012?  
  
SELECT  
  COUNT(*)  
FROM  
  Invoice  
WHERE  
  InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
```

	COUNT(*)
1	167

There were a total of 167 transactions that took place between 2011 and 2012.

```
-- 2. How much money did WSDA Music make during the same period?  
  
SELECT  
  SUM(total)  
FROM  
  Invoice  
WHERE  
  InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
```

	SUM(total)
1	1947.97

Challenge 2

- Get a list of customers who made purchases between 2011 and 2012.
- Get a list of customers, sales reps, and total transaction amounts for each customer between 2011 and 2012.
- How many transactions are above the average transaction amount during the same time period?
- What is the average transaction amount for each year that WSDA Music has been in business?

```

28 -- 3. Get a list of customers who made a purchase between 2011 and 2012.
29
30 SELECT
31     c.FirstName,
32     c.LastName,
33     i.total
34 FROM
35     Invoice AS i
36 INNER JOIN
37     Customer AS c
38 ON
39     c.CustomerId = i.CustomerId
40 WHERE
41     InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
42 ORDER BY
43     i.total DESC
44

```

	FirstName	LastName	Total	
1	John	Doeen	1000.86	
2	Richard	Cunningham	23.86	
3	Hugh	O'Reilly	21.86	
4	Victor	Stevens	18.86	
5	František	Wichterlová	16.86	
6	Isabelle	Mercier	16.86	
7	Bjørn	Hansen	15.86	
8	Fynn	Zimmermann	14.91	
9	Enrique	Muñoz	13.86	
10	Robert	Brown	13.86	
11	Daan	Peeters	13.86	
12	Wyatt	Girard	13.86	
13	Kathy	Chase	13.86	

Execution finished without errors.
Result: 167 rows returned in 23ms

This query returned 167 rows of information relating to the customer's first and last name, and the totals that they've spent between 2011 and 2012.

This query returns a list of all customers, their purchases, and which sales rep fulfilled the purchase.

```
46 -- 4. Get a list of customers, sales reps, and total transaction amounts for each customer between 2011 and 2012.
47
48 SELECT
49     c.FirstName AS 'Customer First Name',
50     c.LastName AS 'Customer Last Name',
51     e.FirstName AS 'Employee First Name',
52     e.LastName AS 'Employee Last Name',
53     i.total
54 FROM
55     Invoice AS i
56 INNER JOIN
57     Customer AS c
58 ON
59     i.CustomerId = c.CustomerId
60 INNER JOIN
61     Employee AS e
62 ON
63     c.SupportRepId = e.EmployeeId
64 WHERE
65     InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
66 ORDER BY
67     i.total DESC
68
```

	Customer First Name	Customer Last Name	Employee First Name	Employee Last Name	Total	
1	John	Doein	Jane	Peacock	1000.86	
2	Richard	Cunningham	Margaret	Park	23.86	
3	Hugh	O'Reilly	Jane	Peacock	21.86	
4	Victor	Stevens	Steve	Johnson	18.86	
5	František	Wichterlová	Margaret	Park	16.86	
6	Isabelle	Mercier	Jane	Peacock	16.86	
7	Bjørn	Hansen	Margaret	Park	15.86	
8	Fynn	Zimmermann	Jane	Peacock	14.91	
9	Enrique	Muñoz	Steve	Johnson	13.86	

Execution finished without errors.

69	-- 5. How many transactions are above the average transaction amount during the same time?
70	
71	-- Find the average transaction amount between 2011 and 2012
72	SELECT
73	round (AVG (total),2) AS 'Average Transaction Amount'
74	FROM
75	Invoice
76	WHERE
77	InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
78	
79	-- Get the number of transactions above the average transaction amount
80	SELECT
81	COUNT (total) AS '# of transactions above avg'
82	FROM
83	Invoice
84	WHERE
85	total >
86	(SELECT
87	round (AVG (total),2) AS 'Average Transaction Amount'
88	FROM
89	Invoice
90	WHERE
91	InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31')
92	AND
93	InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
94	

Average Transaction Amount	
1	11.66

# of transactions above avg	
1	26

Here we execute the first part of our query to find the average transaction amount. Then we nest that query to find the number of transactions that exceeded the average.

```

95 -- 6. What was the average transaction amount for each year that WSDA Music has been in business?
96 SELECT
97     round(AVG(total),2) AS 'Avg Transaction Amount',
98     strftime('%Y', InvoiceDate) AS 'Year'
99 FROM
100     Invoice
101 GROUP BY
102     strftime('%Y', InvoiceDate )
103

```

	Avg Transaction Amount	Year
1	5.42	2009
2	5.8	2010
3	17.51	2011
4	5.75	2012
5	5.63	2013

After calculating the average transaction amount for each year WSDA Music has been in business, we notice a clear outlier in 2011.

Challenge 3

- Get a list of employees who exceeded the average transaction amount from sales they generated during 2011 and 2012.
- Create a Commission Payout column that displays each employee's commission based on 15% of the sales transaction amount.
- Which employee made the highest commission?
- List the customers that the employee identified in the last question.
- Which customer made the highest purchase?
- Is there anything suspicious?
- Who is our primary person of interest?

```

104 -- 7. Get a list of employees who exceeded the average transaction amount from sales they generated during 2011 and 2012.
105 SELECT
106     e.FirstName,
107     e.LastName,
108     sum(i.total) AS 'Total Sales'
109 FROM
110     Invoice AS i
111 INNER JOIN
112     Customer AS c
113 ON
114     i.CustomerId = c.CustomerId
115 INNER JOIN
116     Employee AS e
117 ON
118     e.EmployeeId = c.SupportRepId
119 WHERE
120     InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
121 AND
122     i.total > 11.66
123 GROUP BY
124     e.FirstName,
125     e.LastName
126 ORDER BY
127     e.LastName

```

	FirstName	LastName	Total Sales
1	Steve	Johnson	100.1
2	Margaret	Park	139.74
3	Jane	Peacock	1137.65

Here we calculated which employees exceeded the average transaction amount and Jane Peacock has the most total sales over the average by far.

```

129 -- 8. Create a Commission Payout column that displays each employee's commission based on 15% of the sales transaction amount.
130 SELECT
131     e.FirstName,
132     e.LastName,
133     SUM(i.total) AS 'Total Sales',
134     round(sum(i.total)*.15,2) AS 'Commission Payout'
135 FROM
136     Invoice AS i
137 INNER JOIN
138     Customer AS c
139 ON
140     i.CustomerId = c.CustomerId
141 INNER JOIN
142     Employee AS e
143 ON
144     e.EmployeeId = c.SupportRepId
145 WHERE
146     InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
147 GROUP BY
148     e.FirstName,
149     e.LastName
150 ORDER BY
151     e.LastName

```

	FirstName	LastName	Total Sales	Commission Payout	
1	Steve	Johnson	293.2	43.98	
2	Margaret	Park	322.97	48.45	
3	Jane	Peacock	1331.8	199.77	

We also see that Jane has the most to gain with the 15% commission split to sales agents. Again we see another outlier here.

```

156 -- 10. List the customers that the employee identified in the last question
157 SELECT
158     c.FirstName AS 'Customer First Name',
159     c.LastName AS 'Customer Last Name',
160     e.FirstName AS 'Employee First Name',
161     e.LastName AS 'Employee Last Name',
162     sum(i.total) AS [Total Sales],
163     round(sum(i.total)*.15,2) AS 'Commission Payout'
164 FROM
165     Invoice AS i
166 INNER JOIN
167     Customer AS c
168 ON
169     c.CustomerId = i.CustomerId
170 INNER JOIN
171     employee AS e
172 ON
173     e.EmployeeId = c.SupportRepId
174 WHERE
175     InvoiceDate >= '2011-01-01' AND InvoiceDate <= '2012-12-31'
176 AND
177     c.SupportRepId = '3'
178 GROUP BY
179     c.FirstName,
180     c.LastName,
181     e.FirstName,
182     e.LastName
183 ORDER BY
184     [Total Sales] DESC

```

Now, let's take a look at all of Jane Peacock's customers. In the next slide we'll be able to see the results and if any stand out.

The total sales
from customer
John Doein far
exceed the
others.

	Customer First Name	Customer Last Name	Employee First Name	Employee Last Name	Total Sales	Commission Payout
1	John	Doein	Jane	Peacock	1000.86	150.13
2	Hugh	O'Reilly	Jane	Peacock	32.75	4.91
3	Wyatt	Girard	Jane	Peacock	26.75	4.01
4	Niklas	Schröder	Jane	Peacock	24.75	3.71
5	Puja	Srivastava	Jane	Peacock	24.75	3.71
6	Robert	Brown	Jane	Peacock	24.75	3.71
7	Isabelle	Mercier	Jane	Peacock	19.83	2.97
8	Fynn	Zimmermann	Jane	Peacock	16.89	2.53
9	Terhi	Hämäläinen	Jane	Peacock	16.87	2.53
10	Luís	Gonçalves	Jane	Peacock	16.83	2.52
11	Frank	Ralston	Jane	Peacock	15.88	2.38
12	Edward	Francis	Jane	Peacock	15.84	2.38
13	Tim	Goyer	Jane	Peacock	13.87	2.08
14	Jennifer	Peterson	Jane	Peacock	12.87	1.93
15	Ladislav	Kovács	Jane	Peacock	11.88	1.78
16	Phil	Hughes	Jane	Peacock	11.88	1.78
17	Manoj	Pareek	Jane	Peacock	10.89	1.63
18	Ellie	Sullivan	Jane	Peacock	8.91	1.34
19	Roberto	Almeida	Jane	Peacock	8.91	1.34
20	Emma	Jones	Jane	Peacock	6.93	1.04
21	François	Tremblay	Jane	Peacock	5.94	0.89
22	Michelle	Brooks	Jane	Peacock	2.97	0.45


```

189 --12. Look at this customer record - do you see anything suspicious?
190 SELECT
191     *
192 FROM
193     Customer AS c
194 WHERE
195     c.LastName = 'Doein'
196
197 --Yes, there are no other records pertaining to John Doein. Our Primary person of interest is Jane Peacock.

```

	CustomerId	FirstName	LastName	Company	Address	City	State	Country	PostalCode	Phone	Fax	Email	SupportRepId
1	60	John	Doein		NULL	NULL	NULL	NULL	NULL	NULL	NULL		3

Upon trying to pull more information on John Doein, we notice that there is no other available information about him. We already know that the responsible sales rep is Jane Peacock, who is now our primary suspect in our financial discrepancy.

What people are saying

With this app, I'm
confident to plan a
trip to rural
Vietnam

Wendy Writer, CA

Visual translation
feels like magic

Ronny Reader, NYC

Translate has
officially inspired
me to learn French

Abby Author, NYC