

Corporação Metaphor - Análise e auditoria de dados em um cenário de mundo real.

Esse trabalho visa reproduzir o estudo de caso de uma Empresa fictícia, a *METAPHOR CORPORATION*, disponível no software de auditoria ACL (Audit Command Language). Porém, para as análises, utilizaremos a linguagem *open source* amplamente conhecida, o Python.

Python é uma linguagem de programação de alto nível, interpretada, de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991. Atualmente possui um modelo de desenvolvimento comunitário, aberto e gerenciado pela organização sem fins lucrativos Python Software Foundation.

fonte: <https://pt.wikipedia.org/wiki/Python>

Este tutorial examina registros de recursos humanos e transações com cartão de crédito em uma empresa fictícia, a Metaphor Corporation. Ao trabalhar com os procedimentos e exemplos, você aprenderá a aplicar procedimentos de auditoria e análise de dados em dados do mundo real. Quando terminar, você poderá aplicar o poder e a flexibilidade da linguagem Python em prol da auditoria interna aos seus próprios dados.

Neste tutorial você irá:

- Trabalhar com um projeto de análise de dados com foco em auditoria.
- Criar e trabalhar com um projeto de auditoria de analytics do início ao fim.
- Relatar os resultados do seu projeto.
- Exportar os resultados para outro formato de arquivo.

Essas etapas são típicas de muitos projetos de análise de dados que você executa, sejam eles pequenos ou grandes.

PASSO 1: "Examinar os dados dos funcionários"

Objetivo: Examinar os dados do perfil de funcionários da Metaphor Corporation.

Resumo:

- Examine o projeto e seu conteúdo.
- Criar uma visão geral estatística.
- Examine os bônus e salários dos funcionários.

Sobre a Metaphor Corporation

Desde 1995, a Metaphor vende software e serviços de TI. A Metaphor tem várias dezenas de consultores que são funcionários em tempo integral. Esses consultores trabalham fora da sede da Metaphor e também viajam pelo mundo para instalar software, fornecer cursos de treinamento e iniciar e manter relacionamentos de vendas. Metaphor também tem funcionários administrativos e profissionais na sede. Além disso, a Metaphor possui agentes de vendas em vários locais ao redor do mundo que não são funcionários.

Política de Cartão de Crédito da Metaphor

A maioria dos funcionários possui um cartão de crédito da empresa. Quando um funcionário adquire produtos ou serviços para a empresa, ele deve usar um cartão de crédito da empresa.

Os funcionários entretêm clientes e possíveis clientes, conforme apropriado, na cidade natal de Metaphor e durante a viagem.

Eles também são responsáveis por seu próprio desenvolvimento profissional e devem cobrar as despesas do cartão de crédito da empresa por cursos, livros, assinaturas de periódicos e associações profissionais.

Alguns funcionários também são responsáveis pela compra de suprimentos para o escritório.

Assim, a Metaphor Corporation incorre em despesas significativas em várias categorias, tais como:

- Viagem
- Entretenimento
- Serviço ao cliente
- Desenvolvimento profissional

■ Despesas de escritório

A política da Metaphor é que todas as compras devem ser atribuídas a um número de cliente. Embora a maioria dos números de clientes corresponda a clientes externos, os números de clientes que começam com zero correspondem aos departamentos internos.

Os funcionários estão cientes de que certas categorias de transações são inaceitáveis. Por exemplo, os funcionários não devem usar cartões de crédito corporativos para jogos de azar.

Embora a empresa às vezes aprova transações em outras categorias, a permissão deve ser concedida com antecedência. Algumas categorias são totalmente proibidas.

Ao usar o Python para analisar as transações com cartão de crédito dos funcionários da Metaphor, é possível encontrar transações que sejam contrárias à política e às transações que possam ser errôneas ou fraudulentas.

A tabela a seguir mostra as despesas permitidas:

Viagem	Entreterimento	Escritório e Profissional
Passagem aérea	Restaurantes	Dívidas para clubes de associação
Hotel	Beber lugares	Dívidas para organizações profissionais
Táxi	Barco	Dívidas para organizações fraternas
Restaurantes	fretamento / locação	Livros e periódicos
Carros alugados	Restauração	Computadores, peças de computador e software
Ônibus de longa distância	Padarias	Material de escritório
ingressos	Eventos esportivos	
Agências de viagem	Recreacional	
Taxas de pedágio e ponte	Serviços	
	Parques de diversão	
	Campos de golfe	
	Atrações turísticas	
	Ingressos de teatro	
	Onibus de turismo	

Códigos de categoria aceitáveis.

A análise do cartão de crédito neste tutorial depende da lista de códigos de categoria do comerciante, usada pelos principais fornecedores de cartão de crédito. Ele contém milhares de códigos numéricos de quatro dígitos, incluindo códigos exclusivos para várias grandes companhias aéreas e cadeias de hotéis, além de vários códigos genéricos.

Cada transação com cartão de crédito tem um código de categoria associado atribuído a ela.

No mundo real, as empresas de cartão de crédito ocasionalmente atribuem códigos incorretos, portanto, se você obtiver resultados questionáveis, entre em contato com o comerciante para verificação.

Para os fins desta análise, você pode comparar os códigos de transação às listas de códigos aceitáveis e inaceitáveis. Por exemplo, se um funcionário da Metaphor usa seu cartão de crédito para jogar, a transação código indica uma transação inaceitável.

O documento Políticas, que acompanha os dados desse trabalho, discute as políticas apropriadas de uso do cartão de crédito. O analista de dados usa este documento para determinar quais tipos de transações serão sinalizadas.

Arquivos de dados da Metaphor:

Quando você realiza uma análise para sua própria empresa, geralmente obtém os dados em qualquer formato de arquivo disponível. Neste trabalho, no entanto, você continua um projeto de análise que já foi planejado. Muito do trabalho preliminar já está completo.

Para este projeto de análise, usarei esses arquivos:

■ Acceptable_Codes.xls

Este arquivo em excel tem uma lista de códigos aceitáveis da lista Código de Categoria do Comerciante.

■ Agents_Metaphor.xlsx

Lista todos os agentes de vendas independentes da Metaphor em todo o mundo.

■ Credit_Cards_Metaphor.xls

Esta planilha do Excel lista informações de cartão de crédito.

■ Company_Departments.txt

Este arquivo delimitado por tabulação contém uma lista de departamentos e seus números internos.

■ Employees_List.xlsx

Este arquivo em planilha do excel lista os nomes de funcionários, números de funcionários e números de cartão de crédito.

■ Trans_April.xls

Esta planilha do Excel lista as transações com cartão de crédito para abril de 2003.

■ Unacceptable_Codes.txt

Este arquivo de texto delimitado por tabulação contém uma lista de códigos inaceitáveis da lista Código de Categoria do Comerciante.

Examinando os dados de Funcionários - Analisando salários, bônus e datas de contratação.

Neste capítulo, você observa alguns dados de perfil de funcionários da Metaphor Corporation. Você analisa salários, bônus, datas de contratação e outras informações sobre recursos humanos.

No curso dessa análise, você aprende os fundamentos do Python e como utiliza-lo na análise dos dados.

Nesse cenário, você analisa os dados que seu colega de trabalho reuniu.

Examinando a lista de Funcionários:

Ao analisar os dados dos funcionários, você pode obter algumas ideias sobre esses dados. Neste exemplo, o projeto já foi montado para você.

Quando você trabalha com seus próprios projetos, decide quais dados incluir. Você também pode adicionar mais dados a qualquer momento a um projeto existente.

Sua tarefa é determinar várias informações financeiras e de recursos humanos, com base nos dados do seu projeto.

O projeto com o qual você trabalha neste capítulo representa um projeto de análise de dados em pequena escala, mas bastante típico. Este projeto tem duas tabelas que contêm informações sobre os funcionários da Metaphor. Você executa a análise com base nessas informações.

Obs: Esse trabalho não visa ser um tutorial sobre python, para isso, há um vasto material disponível na internet. Explicarei a função utilizada somente para os casos que julgar necessário.

Nesta parte da análise, usarei duas tabelas: Agents_Metaphor e Employee_List. A tabela Agents_Metaphor lista todos os agentes de vendas independentes da Metaphor em todo o mundo. O Employee_List lista todos os funcionários da Metaphor na matriz.

In [58]:

```
# Importando as biblioteca que serão utilizadas no trabalho.
import pandas as pd

import warnings
warnings.filterwarnings('ignore') # Evitar mensagens de Erros
```

In [2]:

```
# Importando os arquivos para manipulação
Agents_Metaphor = pd.read_excel('Agents_Metaphor.xlsx')
Employee_List = pd.read_excel('Employee_List.xlsx')
```

In [3]:

Visualizando os 5 primeiros registros
Agents_Metaphor.head()

Out[3]:

	ADDRESS	CITY	CODE	COMM_2002	COUNTRY	FIRST1	LAST1	STATE_PROVINCE
0	AV. CENTRAL, CALLES 5 Y 6	ALAJUELA	CR	4220	NaN	CARMEN	BACARDI BOLIVAR	NaN
1	HITDORFER STRASSE 12	ESSEN	DE	3300	NaN	POILIN	BREATNACH	NaN
2	298 MAIN STREET	INDIANAPOLIS	US	3060	UNITED STATES	GREGORY	QUINLAN	IN
3	AACHENER STRASSE 1053-1055	LANGENFELD	DE	3214	NaN	MARKUS	BOCKELKAMP	NaN
4	GROUND FLOOR	COBHAM, SURREY	GB	2580	GREAT BRITAIN	SAVI	MADAN	ENGLAND

In [4]:

Visualizando os 5 primeiros registros
Employee_List.head()

Out[4]:

	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
0	140540	8590122497663807	8	Leila	1997-12-28	Remlawi	52750
1	455743	8590122281964011	60	Vladimir	1997-05-10	Alexov	41250
2	65119	8590120784984566	100	Matthew	1999-03-31	Lee	38250
3	746002	8590124253621744	104	Alex	2001-12-08	Williams	40175
4	699075	8590125999743363	146	Narinder	1999-09-09	Singh	32250

Procurando por Registros Duplicados

Uma maneira de garantir a integridade dos dados é verificar registros duplicados.

Esta pode ser uma tarefa mais complicada do que parece à primeira vista. Um funcionário pode estar listado sob dois nomes ou sob o mesmo nome duas vezes.

Dois funcionários podem ter recebido o mesmo número de funcionário inadvertidamente. O tipo de duplicatas que você deseja verificar depende dos dados com os quais você está trabalhando.

Nos procedimentos a seguir, você verificará se há duplicatas no sobrenome e no número do empregado utilizando a função **duplicated()** da biblioteca Pandas.

In [6]:

```
Last_name_dupl = Employee_List[Employee_List.duplicated('LAST_NAME',keep=False)]# keep = False para manter os dois registros  
# quando duplicados.  
Last_name_dupl.sort_values(by = ['LAST_NAME']) # a função sort_values deixa a coluna LAST_NAME em ordem alfabética crescente.
```

Out[6]:

	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
61	891575	8590128164445837	300025	Wei	1995-07-25	Fisher	31020
155	62394	8590126961897004	800022	Tina	1997-11-28	Fisher	49950
2	65119	8590120784984566	100	Matthew	1999-03-31	Lee	38250
11	892235	8590129593164703	222	James	1998-11-12	Lee	88420
116	975550	8590128263176714	500140	Michal	1997-09-27	Levy	46390
7	183897	8590122720558982	172	Nicole	2002-06-15	Levy	46150
60	628111	8590126611059920	300013	LaVerne	1995-04-26	Michaels	38220
177	915117	8590122716976418	800093	David	1996-07-27	Michaels	47590
38	126185	8590129151565076	200248	Barbara	1996-03-19	Nelson	46250
193	403889	8590128977644129	800233	Greg	1997-01-18	Nelson	43390
20	583179	8590121039256285	328	Beth	1999-11-16	Sinclair	52180

32	500762	8590123888722712	100245	Elisa	1998-12-28	Sinclair	46500
	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
187	524448	8590129875808876	800207	Philip	1999-01-15	Smith	57800
182	809647	8590127452939923	800098	Daniel	2001-05-07	Smith	49300
105	86323	8590125693710607	500103	Connie	1998-02-22	Turner	49280
172	346907	8590124651080991	800079	Bruce	2002-12-06	Turner	59440
48	334967	8590122464207976	200399	Daniel	2001-08-10	Wang	95070
169	46803	8590125699407125	800053	Sean	1998-02-25	Wang	40590

Vemos que apesar de haver sobrenomes duplicados, os primeiros nomes e demais informações são distintas.

Veremos agora se há valores duplicados no campo **EMPNO** (Número do Empregado).

In [7]:

```
Employee_List[Employee_List.duplicated('EMPNO', keep = False)]
```

Out[7]:

BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
------------	---------	-------	------------	----------	-----------	--------

Vemos que não há valores duplicados com relação ao campo EMPNO. A mesma análise pode ser feita com múltiplas colunas, inclusive utilizando-se de todas as colunas.

Encontrar o total de bônus pagos em 2002.

A metaphor tem uma estrutura de compensação complexa. Porque a Metaphor é uma empresa orientada para projetos, alguns funcionários têm um esquema de compensação através de bônus bem alta. Para encontrar o valor total pago em bônus em 2002, podemos utilizar a seguinte linha de comando:

In [8]:

```
Total_Bonus2002 = Employee_List['BONUS_2002'].sum()
print("O valor Total de Bônus pago em 2002 foi de: $',(Total_Bonus2002))
```

O valor Total de Bônus pago em 2002 foi de: \$ 101648376

Verificando os 5 maiores salários pagos em 2002:

In [9]:

```
Employee_List.sort_values('SALARY', ascending= False).head() # A função head() mostra os 5 primeiros registros.
```

Out[9]:

	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
91	864941	8590123081706198	400047	John	1999-11-20	Lawson	98840
52	593423	8590125625945448	200350	Kelly	1997-07-25	Ho	96150
48	334967	8590122464207976	200399	Daniel	2001-08-10	Wang	95070
76	831196	8590126882512521	300143	Fritz	1995-12-11	Zeeman	90080
77	385366	8590125084016948	300189	Leonard	1997-11-09	Kain	88950

O mesmo pode ser aplicado com relação aos maiores bônus pagos em 2002:

In [10]:

```
Employee_List.sort_values('BONUS_2002', ascending= False).head()
```

Out[10]:

BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
------------	---------	-------	------------	----------	-----------	--------

	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
123	994828	8590124910032035	500193	Eveline	1998-06-14	Paris	55370
163	992782	8590127855367963	800038	John	1998-02-18	Geyer	58750
106	992575	8590127958620445	500108	Maria	1995-03-22	Valencia	61250
56	991217	8590121442723858	200362	William	2001-11-11	Samuels	42747

Verificar os Empregados contratados em 2002:

Para determinar quanto empregados foram contratados em 2002, podemos aplicar um filtro dessa maneira:

In [11]:

```
contratados_2002 = Employee_List[(Employee_List['HIREDATE'] > '2002-01-01') & (Employee_List['HIREDATE'] < '2002-12-31')]
print('{} funcionários foram contratados em 2002.'.format(contratados_2002['HIREDATE'].count()))
contratados_2002
```

17 funcionários foram contratados em 2002.

Out[11]:

	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY
5	83698	8590120716753180	157	Albert	2002-09-26	Schmidt	36170
7	183897	8590122720558982	172	Nicole	2002-06-15	Levy	46150
15	197233	8590127188365686	277	Hugh	2002-02-07	Vanda	31340
21	14774	8590122491455345	331	Igor	2002-11-30	Belchev	58760
46	81294	8590129079724782	200345	Rejane	2002-09-09	Boucher	33740
47	602904	8590121278290111	200346	Brett	2002-09-26	Starr	45780
49	919524	8590128375704378	200347	Kim	2002-06-16	Montara	20470
57	948957	8590121785018691	200365	Hilda	2002-07-02	Demeter	45690
63	576533	8590126069373535	300041	Sarah	2002-01-30	Pollard	30200
89	748681	8590121610225371	400044	Brendan	2002-09-20	Hampton	27400
103	4183	8590126272564898	400200	Pedro	2002-10-12	Morales	43600
130	65457	8590124434068810	600018	Sergei	2002-09-13	Zigic	58750
131	603807	8590126821848844	600023	Ruth	2002-11-03	Patterson	46350
148	94254	8590121160951469	800011	Fatima	2002-12-19	Nahan	41400
172	346907	8590124651080991	800079	Bruce	2002-12-06	Turner	59440
174	112921	8590128759809183	800081	Masatoshi	2002-05-09	Yoshimura	56660
175	373810	8590122780127996	800090	Marilyn	2002-06-03	Scoutten	41400

Determinar bônus como porcentagem do salário

Você quer determinar o bônus como uma porcentagem do salário para todos os funcionários, para isso, iremos acrescentar uma coluna com os valores calculados.

In [12]:

```
# Importando os arquivos para manipulação
Agents_Metaphor = pd.read_excel('Agents_Metaphor.xlsx')
Employee_List = pd.read_excel('Employee_List.xlsx')
```

In [13]:

```
# Função para calcular a porcentagem
def percentage(a, b):
    return round(a / b * 100, 2)
```

In [14]:

```
Employee_List['Bonus_Percent_Salary_%'] = percentage(Employee_List['BONUS_2002'], Employee_List['SALARY'])
```

```
Employee_List.sort_values('Bonus_Percent_Salary %').head()
```

Out[14]:

	BONUS_2002	CARDNUM	EMPNO	FIRST_NAME	HIREDATE	LAST_NAME	SALARY	Bonus_Percent_Salary %
103	4183	8590126272564898	400200	Pedro	2002-10-12	Morales	43600	9.59
21	14774	8590122491455345	331	Igor	2002-11-30	Belchev	58760	25.14
90	9332	8590124396138937	400046	Cathy	1998-09-14	Upland	32060	29.11
145	22848	8590125205486541	700102	Eyad	1995-10-25	Sayeed	59440	38.44
118	20305	8590129400665510	500147	Shawna	1996-10-02	Campbell	46380	43.78

Agora entraremos mais profundamente nas análises sobre a Metaphor, para isso, importaremos as bases de dados que serão usadas nas análises a seguir:

O primeiro conjunto de dados a ser importado será o **"Credit_Cards_Metaphor.xls"** que contém as informações sobre os cartões de crédito corporativo.

In [15]:

```
# Importando o dado e converter a coluna "CARDNUM" para texto, desde que não iremos utilizar para nenhuma operação matemática.
Cred_card_metaphor = pd.read_excel('Credit_Cards_Metaphor.xls', dtype = {'CARDNUM': str})
Cred_card_metaphor.head()
```

Out[15]:

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
0	8590122497663807	9000	962353	2005-08-01	NaN	10.0	37.23	0.0	2003-04-28	0.00	12.9	2003-03-29
1	8590122281964011	8000	812465	2004-03-01	14.68	30.0	929.79	10.0	2003-05-03	443.81	12.9	2003-04-03
2	8590120784984566	6900	51593	2004-10-01	113.20	129.0	6408.12	0.0	2003-05-04	6294.92	12.9	2003-04-04
3	8590124253621744	7200	250402	2004-04-01	101.41	0.0	5891.59	0.0	2003-05-14	5850.96	12.9	2003-04-14
4	8590125999743363	9000	778088	2004-02-01	0.00	0.0	384.95	0.0	2003-05-12	1352.51	12.9	2003-04-12

O segundo arquivo a ser importado é o **"Trans_April.xls worksheet"** que contém as transações realizadas no mês de Abril.

In [16]:

```
trans_abril = pd.read_excel('Trans_April.xls')
trans_abril.head()
```

Out[16]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
0	8590120032047834	270.63	2003-04-02	1731	1000	Electrical Contractors
1	8590120092563655	899.76	2003-04-02	1731	2000	Electrical Contractors
2	8590120233319873	730.46	2003-04-04	1750	250402	Carpentry Contractors
3	8590120534914664	106.01	2003-04-08	1750	3000	Carpentry Contractors
4	8590120674263418	309.37	2003-04-08	2741	1000	Miscellaneous Publishing and Printing

O terceiro esta em formato de texto, o **"Unacceptable_Codes.txt"**, que contém os códigos não aceitos para transações com o cartão.

In [17]:

```
# separado por tabulação e o Campo 'Codes' foi convertido para texto, pois apesar de números, ele não será utilizado para nenhuma
# operação aritmética.
cod_naoAceitos = pd.read_csv('Unacceptable_Codes.txt', sep = "\t", dtype = {'Codes': int})
cod_naoAceitos.head()
```

Out[17]:

Codes	Description
-------	-------------

	Codes	Description
0	742	Veterinary Services
1	763	Agriculture Co-operative
2	4011	Railroads
3	4111	Local/Suburban Commuter Passenger Transportati...
4	4411	Cruise Lines

O quarto arquivo a ser importado é o **"Company_Departments.txt"**, que possui informações sobre os departamentos da Cia.

In [18]:

```
dep_cia = pd.read_csv('Company_Departments.txt', sep = "\t", header = None,
names = ['Dept_Name', 'Dept_Code'], dtype = {'Dept_Code': str}) # o parametro 'names =' é aqui utilizado
# para incluir o nome do cabeçalho, ja que o arquivo original não possui.
dep_cia.head()
```

Out[18]:

	Dept_Name	Dept_Code
0	Administrative	000000
1	Product Development	100000
2	Manufacturing	200000
3	Technical Support	300000
4	Sales	400000

O quinto arquivo é o **"Employees.csv"** que contém informação sobre os funcionários. Iremos fazer algumas alterações no tipo dos dados. Iremos converter as colunas "CardNum" e "EmpNo" para texto, "HireDate" para data e "Bonus_2002" para numérico.

In [19]:

```
dtypes = {'CardNum': str, 'EmpNo': str}
employees = pd.read_csv('Employees.csv', sep = ";", dtype = dtypes)
employees['Bonus_2002'] = pd.to_numeric(employees['Bonus_2002'], errors= 'ignore')
employees['HireDate'] = pd.to_datetime(employees['HireDate'])
employees.head()
```

Out[19]:

	First_Name	Last_Name	CardNum	EmpNo	HireDate	Salary	Bonus_2002
0	Leila	Remlawi	8590122497663800	8	1997-12-28	52750	1,405.40
1	Vladimir	Alexov	8590122281964010	60	1997-10-05	41250	4,557.43
2	Matthew	Lee	8590120784984560	100	1999-03-31	38250	651.19
3	Alex	Williams	8590124253621740	104	2001-08-12	40175	7,460.02
4	Narinder	Singh	8590125999743360	146	1999-09-09	32250	6,990.75

O sexto e último arquivo para importar é o **"Acceptable_Codes.xls"**, que contém os códigos aceitos para as transações com os cartões.

In [20]:

```
cod_Aceitos = pd.read_excel('Acceptable_Codes.xls', dtype = {'Codes': str})
cod_Aceitos.head()
```

Out[20]:

	Codes	Description
0	2741	Miscellaneous Publishing and Printing
1	2791	Typesetting, Plate Making, & Related Services
2	3000	United Airlines
3	3001	American Airlines

Codes	Description
4 3002	Pan American

INICIE SUA ANÁLISE: Comece a procurar padrões

Aqui usaremos procedimentos simples para obter informações gerais sobre os dados. Vamos analisar os limites de crédito e os saldos de cartão de crédito para obter informações sobre os padrões de uso de cartão de crédito e responsabilidade financeira da Metaphor. Determinar o passivo total:

Determinar o passivo total:

Se todos os funcionários da Metaphor atingissem o limite de crédito em seu cartão de crédito, qual seria a dívida total da Metaphor? Você pode determinar esse valor totalizando o campo CREDLIM.

Você pode usar o Total para adicionar os limites totais de crédito para determinar a exposição total de crédito da empresa.

In [21]:

```
Cred_card_metaphor.head(3)
```

Out[21]:

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
0	8590122497663807	9000	962353	2005-08-01	NaN	10.0	37.23	0.0	2003-04-28	0.00	12.9	2003-03-29
1	8590122281964011	8000	812465	2004-03-01	14.68	30.0	929.79	10.0	2003-05-03	443.81	12.9	2003-04-03
2	8590120784984566	6900	51593	2004-10-01	113.20	129.0	6408.12	0.0	2003-05-04	6294.92	12.9	2003-04-04

In [22]:

```
print('O valor total de limite de Crédito da Metaphor é $ {:.2f}'.format(Cred_card_metaphor['CREDLIM'].sum(),2))
```

O valor total de limite de Crédito da Metaphor é \$ 1440800.00

Extrair números de cartão expirados

Você planeja coordenar com o departamento de contabilidade para garantir que não haja problemas com os funcionários que recebem novos cartões de crédito. Primeiro, você filtra todos os números de cartão de crédito com data de vencimento de 1º de janeiro de 2004 ou anterior.

In [23]:

```
Cred_card_metaphor[Cred_card_metaphor['EXPDT'] <= '2004-01-01']
```

Out[23]:

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
15	8590127188365686	4400	250402	2004-01-01	54.91	65.0	3098.93	0.0	2003-05-04	3109.02	12.9	2003-04-04
95	8590122098547303	5400	51593	2004-01-01	62.13	64.0	3986.42	0.0	2003-05-03	4004.29	12.9	2003-04-03
105	8590125693710607	9000	778088	2004-01-01	0.00	10.0	14.90	0.0	2003-05-11	0.00	12.9	2003-04-11
109	8590122176454161	9000	51593	2004-01-01	0.00	21.0	963.70	0.0	2003-04-30	822.80	12.9	2003-03-31
154	8590129554875319	1000	250402	2004-01-01	6.90	23.0	398.69	0.0	2003-05-14	414.79	12.9	2003-04-14
155	8590126961897004	5000	444413	2004-01-01	0.00	56.0	2650.44	0.0	2003-05-21	2761.03	12.9	2003-04-21
182	8590127452939923	7000	444413	2004-01-01	0.00	0.0	-0.03	0.0	2003-05-14	-0.03	12.9	2003-04-14
195	8590125918377349	9900	51593	2004-01-01	0.00	13.0	595.12	0.0	2003-05-18	1314.10	12.9	2003-04-18

Como podemos ver, há 8 registros que se encaixam ao nosso critério de pesquisa.

Encontrar transações pelo número do cliente

Por ter sido avisado sobre quaisquer despesas relacionadas ao número de cliente 444413, você deseja encontrar algumas informações sobre as transações associadas a esse número de cliente.

Para isso, procedemos assim (para isso, utilizaremos os dados do conjunto de dados "trans_abril"):

In [24]:

```
trans_abril.head(3)
```

Out[24]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
0	8590120032047834	270.63	2003-04-02	1731	1000	Electrical Contractors
1	8590120092563655	899.76	2003-04-02	1731	2000	Electrical Contractors
2	8590120233319873	730.46	2003-04-04	1750	250402	Carpentry Contractors

In [25]:

```
trans_abril[trans_abril["CUSTNO"]== 444413]
```

Out[25]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
14	8590121039256285	989.14	2003-04-25	3007	444413	Air France
22	8590121410137479	662.22	2003-04-25	3025	444413	Air New Zealand
57	8590123081706198	339.23	2003-04-23	3516	444413	La Quinta Motor Inns
61	8590123423824145	348.60	2003-04-29	3535	444413	Hilton International
71	8590123890848488	359.28	2003-04-06	3535	444413	Hilton International
75	8590124183859322	261.72	2003-04-09	3543	444413	Four Seasons Hotels
84	8590124535028843	986.24	2003-04-19	3649	444413	Radisson Hotels
86	8590124581105605	2.13	2003-04-23	3649	444413	Radisson Hotels
87	8590124651080991	111.00	2003-04-24	3665	444413	Hampton Inns
88	8590124747637528	364.51	2003-04-25	3665	444413	Hampton Inns
100	8590125311193609	20.02	2003-04-14	4121	444413	Taxicabs and Limousines
115	8590126069373535	226.08	2003-04-14	4411	444413	Cruise Lines
137	8590126961897004	745.88	2003-04-07	5094	444413	Precious Stones and Metals, Watches and Jewelry
149	8590127591501905	348.96	2003-04-24	5311	444413	Department Stores
153	8590127875632051	720.29	2003-04-28	5441	444413	Confectionery Stores
154	8590127911877979	175.87	2003-04-28	5441	444413	Confectionery Stores
156	8590127958620445	645.32	2003-04-04	5462	444413	Bakeries
160	8590128152739875	426.19	2003-04-06	5462	444413	Bakeries
183	8590129267949317	148.69	2003-04-04	5811	444413	Caterers
211	8590121278290111	20.78	2003-04-05	5813	444413	Drinking Places (Alcoholic Beverages), Bars, T...
219	8590121837400759	644.57	2003-04-12	5921	444413	Package Stores - Beer, Wine, and Liquor
238	8590125272447003	199.10	2003-04-17	7519	444413	Motor Home and Recreational Vehicle Rentals
243	8590125431232531	654.87	2003-04-23	7933	444413	Bowling Alleys
250	8590128263176714	391.62	2003-04-08	7992	444413	Golf Courses - Public
257	8590128346463420	439.52	2003-04-11	7995	444413	Betting (including Lottery Tickets, Casino Gaming
262	8590121039256285	782.36	2003-04-17	7999	444413	Recreation Services (Not Elsewhere Classified)

Existem 26 registros que atendem ao nosso critério de pesquisa. Em uma análise típica, você provavelmente tentaria descobrir quem é esse cliente. No entanto, os arquivos de dados que acompanham o tutorial não contêm essas informações.

Verificar os saldos dos cartões de crédito.

Você pode obter uma visão geral dos seus dados para alertá-lo sobre possíveis áreas de preocupação. Por exemplo, você pode querer encontrar os saldos mais altos ou encontrar saldos acima de um certo limite. Você pode usar essas informações como um indicador de possíveis problemas com o uso do cartão de crédito do funcionário.

Uma maneira de determinar quantos usuários de cartão de crédito têm saldos elevados seria filtrar os registros para mostrar aqueles com um valor NEWBAL maior que 2000. Ou ordenar em ordem decrescente a coluna NEWBAL para ter uma visão geral dos maiores saldos.

Primeiramente, vamos ordenar em ordem decrescente e apresentar os 15 maiores saldos(Conjunto de dados "Cred_card_metaphor"):

In [26]:

```
Cred_card_metaphor.sort_values('NEWBAL', ascending= False).head(15)
```

Out[26]:

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
123	8590124910032035	9500	250402	2005-05-01	174.76	423.00	10231.00	209.00	2003-05-19	9998.24	12.9	2003-04-19
23	8590129923638380	9000	202028	2004-11-01	73.00	181.00	8671.38	0.00	2003-05-17	6405.43	12.9	2003-04-17
186	8590126772736377	8700	444413	2005-10-01	124.85	180.00	8633.98	0.00	2003-04-28	8691.13	12.9	2003-03-29
53	8590120534914664	6500	202028	2005-03-01	140.34	445.00	7500.24	288.00	2003-05-06	7461.90	12.9	2003-04-06
126	8590125757369024	7100	444413	2004-08-01	110.30	95.00	7045.71	0.00	2003-05-10	7069.43	12.9	2003-04-10
119	8590121044024386	7400	202028	2005-01-01	109.60	41.00	7020.70	0.00	2003-05-03	7017.10	12.9	2003-04-03
131	8590126821848844	7000	812465	2004-06-01	122.00	127.88	6876.02	0.00	2003-05-19	6894.02	12.9	2003-04-19
32	8590123888722712	8200	812465	2004-09-01	95.47	33.00	6580.59	0.00	2003-05-13	6685.12	12.9	2003-04-13
40	8590128906141710	6300	503458	2005-06-01	124.02	233.39	6463.56	98.39	2003-05-03	6581.54	12.9	2003-04-03
2	8590120784984566	6900	51593	2004-10-01	113.20	129.00	6408.12	0.00	2003-05-04	6294.92	12.9	2003-04-04
68	8590124169042633	8000	812465	2005-01-01	88.74	0.00	6262.47	0.00	2003-05-05	5908.17	12.9	2003-04-05
3	8590124253621744	7200	250402	2004-04-01	101.41	0.00	5891.59	0.00	2003-05-14	5850.96	12.9	2003-04-14
29	8590128098022583	5600	962353	2005-02-01	0.00	717.00	5792.57	596.00	2003-05-05	5792.57	12.9	2003-04-05
39	8590128152739875	7000	503458	2005-09-01	0.00	116.00	5543.37	0.00	2003-05-05	-0.22	12.9	2003-04-05
163	8590127855367963	5900	812465	2005-11-01	79.04	0.00	5518.02	0.00	2003-05-07	5439.25	12.9	2003-04-07

Agora filtraremos pelos valores maiores que 2000:

In [27]:

```
Cred_card_metaphor[Cred_card_metaphor['NEWBAL']>= 2000]
```

Out[27]:

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
2	8590120784984566	6900	51593	2004-10-01	113.20	129.00	6408.12	0.00	2003-05-04	6294.92	12.9	2003-04-04
3	8590124253621744	7200	250402	2004-04-01	101.41	0.00	5891.59	0.00	2003-05-14	5850.96	12.9	2003-04-14
15	8590127188365686	4400	250402	2004-01-01	54.91	65.00	3098.93	0.00	2003-05-04	3109.02	12.9	2003-04-04

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	SUMTOT
22	8590127043771382	6500	250402	2003-01-01	87.52	102.00	5060.82	0.00	2003-05-07	4941.37	12.9	5007.07
23	8590129923638380	9000	202028	2004-11-01	73.00	181.00	8671.38	0.00	2003-05-17	6405.43	12.9	2003-04-17
26	8590122965563802	3800	284354	2005-05-01	64.26	0.00	3669.55	0.00	2003-05-18	3705.29	12.9	2003-04-18
28	8590125323000756	4900	503458	2005-06-01	77.35	90.00	4641.63	0.00	2003-04-26	4462.05	12.9	2003-03-27
29	8590128098022583	5600	962353	2005-02-01	0.00	717.00	5792.57	596.00	2003-05-05	5792.57	12.9	2003-04-05
32	8590123888722712	8200	812465	2004-09-01	95.47	33.00	6580.59	0.00	2003-05-13	6685.12	12.9	2003-04-13
34	8590128759193385	9000	202028	2005-11-01	61.19	0.00	4967.92	0.00	2003-04-26	2465.38	12.9	2003-03-27
37	8590121394166062	4900	51593	2005-06-01	66.57	82.00	3921.30	0.00	2003-05-10	3939.73	12.9	2003-04-10
39	8590128152739875	7000	503458	2005-09-01	0.00	116.00	5543.37	0.00	2003-05-05	-0.22	12.9	2003-04-05
40	8590128906141710	6300	503458	2005-06-01	124.02	233.39	6463.56	98.39	2003-05-03	6581.54	12.9	2003-04-03
49	8590128375704378	9000	812465	2004-02-01	48.07	0.00	2773.50	0.00	2003-05-19	2807.01	12.9	2003-04-19
50	8590127538854636	3600	962353	2004-06-01	81.38	662.30	4449.57	569.30	2003-05-11	4318.19	12.9	2003-04-11
53	8590120534914664	6500	202028	2005-03-01	140.34	445.00	7500.24	288.00	2003-05-06	7461.90	12.9	2003-04-06
56	8590121442723858	4300	444413	2004-05-01	74.06	90.00	4285.77	0.00	2003-05-18	4271.00	12.9	2003-04-18
68	8590124169042633	8000	812465	2005-01-01	88.74	0.00	6262.47	0.00	2003-05-05	5908.17	12.9	2003-04-05
69	8590120233319873	4500	812465	2004-10-01	68.64	56.00	2657.96	0.00	2003-05-20	2447.30	12.9	2003-04-20
71	8590121747771148	8600	51593	2005-10-01	58.21	0.00	3146.66	0.00	2003-05-20	3088.45	12.9	2003-04-20
72	8590128540908810	6500	778088	2005-10-01	70.94	0.00	4078.94	0.00	2003-04-26	4208.00	12.9	2003-03-27
73	8590123890848488	8900	778088	2005-08-01	0.00	43.00	2041.80	0.00	2003-04-28	59.93	12.9	2003-03-29
74	8590124466089977	8600	250402	2005-04-01	51.08	0.00	3039.61	0.00	2003-05-18	2980.63	12.9	2003-04-18
80	8590120771950301	6500	202028	2005-09-01	46.13	103.00	4941.13	0.00	2003-04-29	5000.00	12.9	2003-03-30
88	8590125273861859	9000	878035	2004-03-01	65.49	77.00	3656.70	0.00	2003-05-05	3668.21	12.9	2003-04-05
95	8590122098547303	5400	51593	2004-01-01	62.13	64.00	3986.42	0.00	2003-05-03	4004.29	12.9	2003-04-03
108	8590126874849092	2400	925007	2005-03-01	49.06	184.00	2977.67	121.00	2003-05-05	2893.61	12.9	2003-04-05
111	8590125281347499	7400	812465	2005-08-01	67.82	0.00	3961.14	0.00	2003-05-19	3993.32	12.9	2003-04-19
117	8590128006917664	7900	925007	2004-04-01	84.82	0.00	4903.49	0.00	2003-05-21	4901.30	12.9	2003-04-21
119	8590121044024386	7400	202028	2005-01-01	109.60	41.00	7020.70	0.00	2003-05-03	7017.10	12.9	2003-04-03
123	8590124910032035	9500	250402	2005-05-01	174.76	423.00	10231.00	209.00	2003-05-19	9998.24	12.9	2003-04-19
124	8590120923083354	7400	51593	2005-02-01	66.53	53.00	3889.76	0.00	2003-04-28	3924.28	12.9	2003-03-29
125	8590124581105605	7500	812465	2004-03-01	55.51	0.00	2996.64	0.00	2003-04-29	2250.86	12.9	2003-03-30
126	8590125757369024	7100	444413	2004-08-01	110.30	95.00	7045.71	0.00	2003-05-10	7069.43	12.9	2003-04-10
127	8590120772096763	9000	250402	2005-01-01	0.00	365.00	3468.81	292.00	2003-05-21	3468.81	12.9	2003-04-21

	CARDNUM	CREDLIM	CUSTNO	EXPDY	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
131	8590126821848844	7000	812465	2004-06-01	122.00	127.88	6876.02	0.00	2003-05-19	6894.02	12.9	2003-04-19
133	8590120969360742	1500	962353	2004-05-01	48.84	57.00	2735.73	0.00	2003-05-14	2726.89	12.9	2003-04-14
136	85901211121460384	2200	284354	2005-07-01	50.52	59.00	2795.27	0.00	2003-04-29	2805.75	12.9	2003-03-30
140	8590120807946744	9000	51593	2004-07-01	0.00	44.00	2066.29	0.00	2003-05-07	848.61	12.9	2003-04-07
142	8590127734155216	2000	202028	2005-01-01	36.76	230.00	2240.06	184.00	2003-05-05	2168.30	12.9	2003-04-05
145	8590125205486541	6400	284354	2005-11-01	73.21	178.00	4312.16	88.00	2003-04-30	4209.95	12.9	2003-03-31
148	85901211160951469	5000	202028	2004-04-01	57.86	400.00	3485.68	300.00	2003-05-13	3412.82	12.9	2003-04-13
149	8590129792699147	4200	51593	2004-08-01	49.54	40.00	2806.15	0.00	2003-05-17	2831.61	12.9	2003-04-17
155	8590126961897004	5000	444413	2004-01-01	0.00	56.00	2650.44	0.00	2003-05-21	2761.03	12.9	2003-04-21
156	8590128601426195	8000	812465	2004-03-01	31.11	0.00	2257.96	0.00	2003-05-20	2040.88	12.9	2003-04-20
158	8590125311193609	8000	778088	2005-09-01	52.54	0.00	2235.97	0.00	2003-04-28	1068.82	12.9	2003-03-29
161	8590129982969758	7000	925007	2005-07-01	103.27	0.00	4287.33	0.00	2003-05-17	7542.06	12.9	2003-04-17
163	8590127855367963	5900	812465	2005-11-01	79.04	0.00	5518.02	0.00	2003-05-07	5439.25	12.9	2003-04-07
176	8590125589521771	4900	284354	2005-01-01	44.88	113.00	2690.60	56.00	2003-05-12	2680.72	12.9	2003-04-12
177	8590122716976418	9000	284354	2004-10-01	79.31	0.00	5070.35	0.00	2003-05-17	3768.61	12.9	2003-04-17
185	8590126382371244	6400	444413	2005-02-01	58.14	132.00	3171.82	65.00	2003-04-27	3084.68	12.9	2003-03-28
186	8590126772736377	8700	444413	2005-10-01	124.85	180.00	8633.98	0.00	2003-04-28	8691.13	12.9	2003-03-29
198	8590122974573538	4900	51593	2004-07-01	0.00	136.00	3790.92	0.00	2003-05-21	4062.92	12.9	2003-04-21

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
127	8590120772096763	9000	250402	2005-01-01	0.00	365.00	3468.81	292.00	2003-05-21	3468.81	12.9	2003-04-21
142	8590127734155216	2000	202028	2005-01-01	36.76	230.00	2240.06	184.00	2003-05-05	2168.30	12.9	2003-04-05
145	8590125205486541	6400	284354	2005-11-01	73.21	178.00	4312.16	88.00	2003-04-30	4209.95	12.9	2003-03-31
148	8590121160951469	5000	202028	2004-04-01	57.86	400.00	3485.68	300.00	2003-05-13	3412.82	12.9	2003-04-13
176	8590125589521771	4900	284354	2005-01-01	44.88	113.00	2690.60	56.00	2003-05-12	2680.72	12.9	2003-04-12
185	8590126382371244	6400	444413	2005-02-01	58.14	132.00	3171.82	65.00	2003-04-27	3084.68	12.9	2003-03-28

Temos 12 registros que atendem a ambos os critérios.

Agora vamos verificar quais funcionários tem valores vencidos e mostra-los em forma decrescente:

In [29]:

```
Cred_card_metaphor[Cred_card_metaphor["PASTDUEAMT"] > 0 ].sort_values("PASTDUEAMT", ascending=False)
```

Out[29]:

	CARDNUM	CREDLIM	CUSTNO	EXPDT	FINCHG	MINPYMTDUE	NEWBAL	PASTDUEAMT	PMTDUEDT	PREVBAL	RATE	STMTDT
29	8590128098022583	5600	962353	2005-02-01	0.00	717.00	5792.57	596.00	2003-05-05	5792.57	12.9	2003-04-05
50	8590127538854636	3600	962353	2004-06-01	81.38	662.30	4449.57	569.30	2003-05-11	4318.19	12.9	2003-04-11
148	8590121160951469	5000	202028	2004-04-01	57.86	400.00	3485.68	300.00	2003-05-13	3412.82	12.9	2003-04-13
127	8590120772096763	9000	250402	2005-01-01	0.00	365.00	3468.81	292.00	2003-05-21	3468.81	12.9	2003-04-21
53	8590120534914664	6500	202028	2005-03-01	140.34	445.00	7500.24	288.00	2003-05-06	7461.90	12.9	2003-04-06
123	8590124910032035	9500	250402	2005-05-01	174.76	423.00	10231.00	209.00	2003-05-19	9998.24	12.9	2003-04-19
142	8590127734155216	2000	202028	2005-01-01	36.76	230.00	2240.06	184.00	2003-05-05	2168.30	12.9	2003-04-05
17	8590125431232531	1300	503458	2004-07-01	34.76	225.00	1986.63	183.00	2003-04-30	1893.87	12.9	2003-03-31
108	8590126874849092	2400	925007	2005-03-01	49.06	184.00	2977.67	121.00	2003-05-05	2893.61	12.9	2003-04-05
40	8590128906141710	6300	503458	2005-06-01	124.02	233.39	6463.56	98.39	2003-05-03	6581.54	12.9	2003-04-03
145	8590125205486541	6400	284354	2005-11-01	73.21	178.00	4312.16	88.00	2003-04-30	4209.95	12.9	2003-03-31
153	8590123361284174	1700	812465	2005-09-01	0.00	102.00	1562.18	68.00	2003-04-29	1562.18	12.9	2003-03-30
185	8590126382371244	6400	444413	2005-02-01	58.14	132.00	3171.82	65.00	2003-04-27	3084.68	12.9	2003-03-28
176	8590125589521771	4900	284354	2005-01-01	44.88	113.00	2690.60	56.00	2003-05-12	2680.72	12.9	2003-04-12
5	8590120716753180	8000	778088	2005-01-01	0.50	20.79	85.20	10.79	2003-05-07	10.79	12.9	2003-04-07
1	8590122281964011	8000	812465	2004-03-01	14.68	30.00	929.79	10.00	2003-05-03	443.81	12.9	2003-04-03
62	8590126256532094	7000	202028	2004-11-01	9.90	21.00	519.27	10.00	2003-05-11	426.37	12.9	2003-04-11
58	8590125574108273	8000	962353	2004-07-01	2.74	20.00	140.34	10.00	2003-04-29	108.60	12.9	2003-03-30
31	8590128742834790	9500	284354	2004-05-01	1.53	20.00	89.19	10.00	2003-04-29	45.76	12.9	2003-03-30
93	8590120092563655	6800	202028	2004-11-01	8.07	17.52	383.22	7.52	2003-04-30	229.32	12.9	2003-03-31

Temos 20 registros que atendem ao critério estipulado.

Aqui vemos que há inúmeras oportunidades de análises distintas, basta conhecer bem o negócio e os conjuntos de dados analisados para tirar vários insights significantes para posteriores análises e coleta de novas evidências!

EXAMINE PADRÕES DE DESPESAS: Determine as informações sobre as despesas

Até agora trabalhamos com o Python para obter familiaridade com os dados e realizar uma análise preliminar dos dados. Agora, o departamento de contabilidade da Metaphor Corporation quer que você forneça informações sobre padrões de gastos nas transações com cartão de crédito de abril. Se você puder desenvolver um plano de análise útil, será solicitado que você realize análises para outros meses.

Depois que você fornecer informações sobre os padrões de gastos, a Metaphor poderá economizar escolhendo um único fornecedor em alguns casos, ou obtendo descontos por atacado, ou limitando ainda mais as categorias aceitáveis de transações. Você também analisará as despesas por cliente para que a Metaphor possa usar os resultados para determinar se o dinheiro está sendo gasto de forma eficaz.

Sumarizar as despesas totais para cada categoria.

Você inicia sua análise resumindo o valor total de todas as transações para cada código de categoria.

Seu plano é analisar esses resultados para obter ideias para mais análises.

Resumindo os valores de transação para cada código de categoria de comerciante, você pode ver quanto dinheiro foi gasto em determinadas companhias aéreas e cadeias de hotéis específicas.

Todas as despesas do restaurante, por outro lado, serão categorizadas em um único código. Alguns códigos de categoria do comerciante são específicos, enquanto outros são amplos.

Para isso utilizaremos a planilha trans_abril.

In [30]:

```
trans_abril.head()
```

Out[30]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
0	8590120032047834	270.63	2003-04-02	1731	1000	Electrical Contractors
1	8590120092563655	899.76	2003-04-02	1731	2000	Electrical Contractors
2	8590120233319873	730.46	2003-04-04	1750	250402	Carpentry Contractors
3	8590120534914664	106.01	2003-04-08	1750	3000	Carpentry Contractors
4	8590120674263418	309.37	2003-04-08	2741	1000	Miscellaneous Publishing and Printing

Iremos agrupar pelo campo "CODES" e "DESCRIPTION" e mostraremos os 10 maiores valores totais(Campo "AMOUNT"):

In [31]:

```
sumario_abril_codes = trans_abril.groupby(['CODES','DESCRIPTION'],
                                           as_index=False)['AMOUNT'].sum().sort_values('AMOUNT',ascending = False)
sumario_abril_codes.head(10)
```

Out[31]:

	CODES	DESCRIPTION	AMOUNT
67	5811	Caterers	9023.74
68	5812	Eating places and Restaurants	7626.05
26	3535	Hilton International	6043.64
64	5462	Bakeries	4764.13
90	7941	Commercial Sports, Athletic Fields, Profession...	3474.76
92	7992	Golf Courses - Public	3330.45
101	8699	Professional Services (Not Elsewhere Defined)	3187.71

24	CODES	DESCRIPTION	AMOUNT
	3516	La Quinta Motor Inns	3179.86
61	5192	Books, Periodicals, and Newspapers	2600.50
15	3066	Southwest	2455.74

Essa tabela de resumo é útil para mostrar os valores gastos em companhias aéreas, cadeias de hotéis, agências de aluguel de carros e outras categorias. No entanto, não mostra o montante total gasto em cada uma das passagens aéreas, hotéis e aluguel de carros, nem informa o montante total gasto em viagens.

Muitas vezes, quando você executa uma tarefa analítica pela primeira vez, não obtém precisamente os resultados desejados. No entanto, você pode usar esses resultados para ajudá-lo a encontrar uma maneira de alcançar resultados mais focados.

Isolando Transações por Categorias:

A Metaphor quer que você analise de perto o valor total gasto em estadias fora da cidade, em abril, especificamente em passagens aéreas, acomodação e aluguel de carros.

Hotéis, motéis, aluguel de carros e passagem aérea caem exclusivamente no intervalo de códigos 3000-3750. Todas as transações com esses códigos devem estar relacionadas a viagens de funcionários da Metaphor.

Você pode isolar essas despesas filtrando e exibindo todas as transações com códigos entre e incluindo de 3000 a 3750.

In [32]:

```
trans_abril.head()
```

Out[32]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
0	8590120032047834	270.63	2003-04-02	1731	1000	Electrical Contractors
1	8590120092563655	899.76	2003-04-02	1731	2000	Electrical Contractors
2	8590120233319873	730.46	2003-04-04	1750	250402	Carpentry Contractors
3	8590120534914664	106.01	2003-04-08	1750	3000	Carpentry Contractors
4	8590120674263418	309.37	2003-04-08	2741	1000	Miscellaneous Publishing and Printing

In [33]:

```
trans_abril_viagens = trans_abril[(trans_abril['CODES'] >=3000) & (trans_abril['CODES'] <= 3750)]
trans_abril_viagens
```

Out[33]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
8	8590120784984566	555.71	2003-04-07	3000	925007	United Airlines
9	8590120807946744	842.96	2003-04-17	3000	878035	United Airlines
10	8590120860716211	921.27	2003-04-09	3001	812465	American Airlines
11	8590120884250381	44.22	2003-04-10	3001	778088	American Airlines
12	8590120923083354	475.31	2003-04-12	3005	284354	British Airways
13	8590120969360742	197.62	2003-04-18	3005	359310	British Airways
14	8590121039256285	989.14	2003-04-25	3007	444413	Air France
15	8590121044024386	595.92	2003-04-28	3007	309821	Air France
16	8590121105619322	707.19	2003-04-07	3008	980918	Lufthansa
17	8590121121460384	553.17	2003-04-28	3008	51575	Lufthansa
18	8590121278290111	937.51	2003-04-21	3008	203018	Lufthansa
19	8590121282195395	533.22	2003-04-22	3009	290202	Air Canada
20	8590121300586153	920.23	2003-04-24	3009	284354	Air Canada
21	8590121394166062	839.17	2003-04-01	3009	359701	Air Canada
22	8590121410137479	662.22	2003-04-25	3025	444413	Air New Zealand
23	8590121442723858	867.39	2003-04-27	3025	503458	Air New Zealand

24	CARDNUM 8590121469799624	AMOUNT 806.89	TRANS_DATE 2003-04-29	CODES 3025	CUSTNO 962353	DESCRIPTION Air New Zealand
25	8590121610225371	630.72	2003-04-30	3032	925007	El Al
26	8590121620630401	540.95	2003-04-09	3032	878035	El Al
27	8590121747771148	832.87	2003-04-11	3032	878035	El Al
28	8590121762084715	826.63	2003-04-13	3058	878035	Delta
29	8590121785018691	511.81	2003-04-04	3058	812465	Delta
30	8590121791180186	725.83	2003-04-15	3058	202028	Delta
31	8590121807674351	597.67	2003-04-15	3060	250402	Northwest
32	8590121837400759	885.94	2003-04-13	3060	202028	Northwest
33	8590121873994032	29.68	2003-04-16	3063	925007	Us Air
34	8590122098547303	710.36	2003-04-03	3063	962353	Us Air
35	8590122176454161	689.36	2003-04-16	3066	284354	Southwest
36	8590122194440387	408.61	2003-04-17	3066	503458	Southwest
37	8590122221251495	211.45	2003-04-19	3066	503458	Southwest
...
66	8590123730566134	243.15	2003-04-02	3535	51575	Hilton International
67	8590123855538829	219.43	2003-04-06	3535	203018	Hilton International
68	8590123856474124	201.46	2003-04-03	3535	290202	Hilton International
69	8590123870802696	717.21	2003-02-02	3535	203018	Hilton International
70	8590123888722712	882.21	2003-04-16	3535	290202	Hilton International
71	8590123890848488	359.28	2003-04-06	3535	444413	Hilton International
72	8590123916128793	692.11	2003-04-19	3535	778088	Hilton International
73	8590124072645078	438.09	2003-04-08	3543	284354	Four Seasons Hotels
74	8590124156904525	688.12	2003-04-05	3543	359310	Four Seasons Hotels
75	8590124183859322	261.72	2003-04-09	3543	444413	Four Seasons Hotels
76	8590124214758062	518.59	2003-04-10	3543	309821	Four Seasons Hotels
77	8590124238361165	258.77	2003-04-11	3543	503458	Four Seasons Hotels
78	8590124253621744	630.33	2003-04-11	3637	51575	Ramada Inns
79	8590124265829051	637.13	2003-04-13	3638	51593	Howard Johnson
80	8590124396138937	509.05	2003-04-04	3638	812465	Howard Johnson
81	8590124434068810	619.93	2003-04-15	3640	812465	Hyatt Hotel
82	8590124466089977	542.53	2003-04-15	3640	284354	Hyatt Hotel
83	8590124516215636	411.44	2003-04-17	3640	812465	Hyatt Hotel
84	8590124535028843	986.24	2003-04-19	3649	444413	Radisson Hotels
85	8590124558281199	49.36	2003-04-21	3649	503458	Radisson Hotels
86	8590124581105605	2.13	2003-04-23	3649	444413	Radisson Hotels
87	8590124651080991	111.00	2003-04-24	3665	444413	Hampton Inns
88	8590124747637528	364.51	2003-04-25	3665	444413	Hampton Inns
89	8590124781270125	38.94	2003-04-26	3700	812465	Motel 6
90	8590124910032035	469.00	2003-04-27	3703	250402	Residence Inns
91	8590125049097710	260.49	2003-04-27	3730	359310	Mgm Grand Hotel
92	8590125071756786	456.13	2003-04-13	3731	359310	Harrah's Hotels And Casinos
93	8590125084016948	580.95	2003-04-14	3737	359310	Riviera Hotel And Casino
94	8590125164950297	750.12	2003-04-09	3742	202028	Club Med
95	8590125205486541	113.57	2003-01-04	3750	51593	Crowne Plaza Hotels

88 rows × 6 columns

Agora fica fácil sabermos o total gasto com viagens em Abril, para isso:

In [34]:

```
print('O total gasto com viagens em abril foi de $',trans_abril_viagens['AMOUNT'].sum())
```

O total gasto com viagens em abril foi de \$ 47248.27

Se quisermos filtrar apenas pelas passagens aéreas, é só filtramos pelos códigos entre 3000 e 3299:

In [35]:

```
trans_abril_passagens_aereas = trans_abril[(trans_abril['CODES'] >=3000) & (trans_abril['CODES'] <= 3299)]
trans_abril_passagens_aereas
```

Out[35]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
8	8590120784984566	555.71	2003-04-07	3000	925007	United Airlines
9	8590120807946744	842.96	2003-04-17	3000	878035	United Airlines
10	8590120860716211	921.27	2003-04-09	3001	812465	American Airlines
11	8590120884250381	44.22	2003-04-10	3001	778088	American Airlines
12	8590120923083354	475.31	2003-04-12	3005	284354	British Airways
13	8590120969360742	197.62	2003-04-18	3005	359310	British Airways
14	8590121039256285	989.14	2003-04-25	3007	444413	Air France
15	8590121044024386	595.92	2003-04-28	3007	309821	Air France
16	8590121105619322	707.19	2003-04-07	3008	980918	Lufthansa
17	8590121121460384	553.17	2003-04-28	3008	51575	Lufthansa
18	8590121278290111	937.51	2003-04-21	3008	203018	Lufthansa
19	8590121282195395	533.22	2003-04-22	3009	290202	Air Canada
20	8590121300586153	920.23	2003-04-24	3009	284354	Air Canada
21	8590121394166062	839.17	2003-04-01	3009	359701	Air Canada
22	8590121410137479	662.22	2003-04-25	3025	444413	Air New Zealand
23	8590121442723858	867.39	2003-04-27	3025	503458	Air New Zealand
24	8590121469799624	806.89	2003-04-29	3025	962353	Air New Zealand
25	8590121610225371	630.72	2003-04-30	3032	925007	EI AI
26	8590121620630401	540.95	2003-04-09	3032	878035	EI AI
27	8590121747771148	832.87	2003-04-11	3032	878035	EI AI
28	8590121762084715	826.63	2003-04-13	3058	878035	Delta
29	8590121785018691	511.81	2003-04-04	3058	812465	Delta
30	8590121791180186	725.83	2003-04-15	3058	202028	Delta
31	8590121807674351	597.67	2003-04-15	3060	250402	Northwest
32	8590121837400759	885.94	2003-04-13	3060	202028	Northwest
33	8590121873994032	29.68	2003-04-16	3063	925007	Us Air
34	8590122098547303	710.36	2003-04-03	3063	962353	Us Air
35	8590122176454161	689.36	2003-04-16	3066	284354	Southwest
36	8590122194440387	408.61	2003-04-17	3066	503458	Southwest
37	8590122221251495	211.45	2003-04-19	3066	503458	Southwest
38	8590122261490788	866.59	2003-04-20	3066	812465	Southwest
39	8590122281964011	19.24	2003-04-28	3066	778088	Southwest
40	8590122381855730	194.16	2003-04-23	3066	284354	Southwest
41	8590122464207976	66.33	2003-04-01	3066	503458	Southwest
42	8590122491455345	589.77	2003-04-26	3085	962353	Midwest Express Airlines, Inc.
43	8590122497663807	365.07	2003-04-28	3256	812465	Alaska Airlines
44	8590122548770459	539.92	2003-04-29	3261	878035	Air China

In [36]:

```
print('O total gasto com passagens aéreas em abril foi de $',trans_abril_passagens_aereas['AMOUNT'].sum())
```

O total gasto com passagens aéreas em abril foi de \$ 21692.1

Isolar despesas recreativas específicas

Até agora, você examinou as despesas relacionadas a viagens em abril.

A Metaphor quer que você veja as despesas totais relacionadas aos estabelecimentos de golfe e bebidas. Você precisa isolar essas transações com um código de categoria de comerciante relacionado a estabelecimentos de golfe ou bebidas. 5813 é o código para estabelecimentos de bebidas e 7992 para campos de golfe.

In [37]:

```
trans_abril_golfe_bebidas = trans_abril[(trans_abril['CODES'] == 5813) |
                                         (trans_abril['CODES'] == 7992)]

trans_abril_golfe_bebidas
```

Out[37]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
209	8590121121460384	2.06	2003-04-04	5813	778088	Drinking Places (Alcoholic Beverages), Bars, T...
210	8590121160951469	460.63	2003-04-04	5813	359310	Drinking Places (Alcoholic Beverages), Bars, T...
211	8590121278290111	20.78	2003-04-05	5813	444413	Drinking Places (Alcoholic Beverages), Bars, T...
212	8590121282195395	311.64	2003-04-17	5813	202028	Drinking Places (Alcoholic Beverages), Bars, T...
213	8590121300586153	830.32	2003-04-08	5813	778088	Drinking Places (Alcoholic Beverages), Bars, T...
250	8590128263176714	391.62	2003-04-08	7992	444413	Golf Courses - Public
251	8590128098022583	306.95	2003-04-09	7992	151583	Golf Courses - Public
252	8590128152739875	785.00	2003-04-09	7992	309821	Golf Courses - Public
253	8590128164445837	674.04	2003-04-10	7992	203018	Golf Courses - Public
254	8590128221143124	916.92	2003-04-04	7992	290202	Golf Courses - Public
255	8590128263176714	255.92	2003-04-12	7992	284354	Golf Courses - Public

In [38]:

```
print('O total gasto com despesas recreativas(golfe e bebidas) foi de $',trans_abril_golfe_bebidas['AMOUNT'].sum())
```

O total gasto com despesas recreativas(golfe e bebidas) foi de \$ 4955.88

Listar os Totais gastos para cada número de cartão de crédito

Agora que você tem uma lista completa de transações, você deseja determinar o total de valores de transação para cada número de cartão de crédito referente ao mês de abril. Para isso, podemos utilizar a função groupby() do pandas.

In [39]:

```
total_por_cartao = trans_abril.groupby('CARDNUM', as_index=False)['AMOUNT'].sum()
total_por_cartao.sort_values('AMOUNT',ascending = False).head(10) # Apenas os 10 primeiros casos ordenados
#por ordem decrescente de valor.
```

Out[39]:

	CARDNUM	AMOUNT
15	8590121044024386	3130.40

	CARD_NUM	AMOUNT
14	8590121105619322	2450.11
79	8590124253621744	2073.42
47	8590122631528386	1810.97
21	8590121300586153	1750.55
55	8590122974573538	1683.82
39	8590122261490788	1661.36
22	8590121394166062	1656.93
191	8590129923638380	1635.44

ANALISAR TRANSAÇÕES: Identificar transações inaceitáveis

Até agora realizamos uma investigação preliminar e análise dos dados da Metaphor. Agora vamos determinar se os funcionários da Metaphor estão usando os cartões de crédito da empresa para fazer compras não aprovadas e se estão usando os cartões de crédito da empresa para comprar itens e serviços especificamente proibidos.

Determinar quais transações são inaceitáveis.

A Metaphor Corporation criou três categorias de códigos de transação, com base na lista de códigos de categoria do comerciante. Esses incluem:

- Aceitável
- Inaceitável
- Condicional (requer aprovação prévia)

Uma tabela lista os códigos aceitáveis e uma tabela lista os códigos inaceitáveis. Códigos que não estão listados em nenhuma tabela são condicionais.

Sobre os comando merge()

Quando seu projeto inclui várias tabelas de dados, você geralmente quer descobrir se existem valores de campos específicos em mais de uma tabela. Exemplos:

- Se você gerencia uma agência de viagens, você pode ter uma lista de clientes que compraram passagens aéreas e uma lista de clientes que compraram quartos de hotel. Você pode determinar quantos clientes compraram tarifas aéreas e quartos de hotel da sua agência.
- Se você trabalha em marketing em uma loja de departamento, pode ter uma lista de clientes que compraram mais de US 2000 no ano passado em sua loja e uma lista de clientes que obtiveram reembolsos de mais de US 1.000. Você pode determinar quais clientes compraram mais de US 2000 no ano passado e também obtiveram um reembolso de mais de US 1.000.

Você pode usar os comandos Join e Relations para comparar tabelas e combinar dados de tabelas. Com esses comandos, você pode:

- Encontrar valores de campo que existem em uma tabela, mas não em outra.
- Encontre valores de campos que existem nas duas tabelas.
- Combinar campos especificados de duas tabelas de origem. O comando merge() cria uma nova tabela que contém dados de ambas as tabelas originais. A tabela resultante é semelhante a qualquer outra tabela de dados em seu projeto.

Verifique as tabelas de códigos em busca de erros óbvios

Antes de poder usar as tabelas de códigos para verificar as transações, verifique as tabelas em busca de erros óbvios. Você primeiramente tem que determinar se a tabela Acceptable_Codes e a tabela Unacceptable_Codes possuem códigos em comum (Seria um erro grave se houver, portanto esperamos que o resultado seja 0).

Vamos verificar agora como proceder para atingir esse objetivo usando o Python:

In [40]:

```
cod_Aceitos.head()
```

Out[40]:

	Codes	Description
0	2741	Miscellaneous Publishing and Printing
		Transaction, Plate Making & Related

1	Codes	Description
2	3000	United Airlines
3	3001	American Airlines
4	3002	Pan American

In [41]:

```
cod_ naoAceitos.head()
```

Out[41]:

	Codes	Description
0	742	Veterinary Services
1	763	Agriculture Co-operative
2	4011	Railroads
3	4111	Local/Suburban Commuter Passenger Transportati...
4	4411	Cruise Lines

In [42]:

```
comparacao = [ num for cod in cod_Aceitos['Codes'] if cod in cod_ naoAceitos['Codes']]
comparacao
```

Out[42]:

```
[]
```

O código acima faz uma comparação de cada código('Codes') do conjunto de dados 'cod_Aceitos' com cada código('Codes') do conjunto de dados 'cod_ naoAceitos'. Como vimos, retornou uma lista vazia, o que significa que não há valores iguais em ambos os conjuntos de dados.

Identificar transações inaceitáveis

A tabela Trans_April contém estes campos:

- CARDNUM
- AMOUNT
- TRANS_DATE
- CODES
- CUSTNO
- DESCRIPTION

A tabela Unacceptable_Codes contém o seguinte Campos:

- Codes
- Description
- Codes_Sub_Unacceptable (a computed field)

Para identificar transações inaceitáveis, você precisa relacionar a tabela Trans_April com a tabela Unacceptable_Codes. Para isso devemos verificar quais campos em comum há em ambas tabelas para podermos relacionar os dados.

In [43]:

```
trans_abril.head()
```

Out[43]:

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
0	8590120032047834	270.63	2003-04-02	1731	1000	Electrical Contractors
1	8590120092563655	899.76	2003-04-02	1731	2000	Electrical Contractors

	CARDNUM	AMOUNT	TRANS_DATE	CODES	CUSTNO	DESCRIPTION
2	8590120233319873	730.46	2003-04-04	1750	250402	Carpentry Contractors
3	8590120534914664	106.01	2003-04-08	1750	3000	Carpentry Contractors
4	8590120674263418	309.37	2003-04-08	2741	1000	Miscellaneous Publishing and Printing

In [44]:

```
cod_naoAceitos.head()
```

Out[44]:

	Codes	Description
0	742	Veterinary Services
1	763	Agriculture Co-operative
2	4011	Railroads
3	4111	Local/Suburban Commuter Passenger Transportati...
4	4411	Cruise Lines

In [45]:

```
cod_naoAceitos = cod_naoAceitos.rename(columns={'Codes':'CODES'}) # Deixando o nome igual ao da tabela trans_abril
```

In [46]:

```
trans_abril_naoAceitos = pd.merge(cod_naoAceitos,trans_abril,
                                  on= 'CODES', how='inner')
trans_abril_naoAceitos.drop('Description', axis = 1, inplace = True)# Para evitar a duplicidade da coluna Description.
```

In [47]:

```
trans_abril_naoAceitos
```

Out[47]:

	CODES	CARDNUM	AMOUNT	TRANS_DATE	CUSTNO	DESCRIPTION
0	4411	8590126009071523	314.33	2003-04-14	284354	Cruise Lines
1	4411	8590126069373535	226.08	2003-04-14	444413	Cruise Lines
2	5094	8590126961897004	745.88	2003-04-07	444413	Precious Stones and Metals, Watches and Jewelry
3	5094	8590127015016937	457.70	2003-04-09	962353	Precious Stones and Metals, Watches and Jewelry
4	7994	8590128345724706	642.83	2003-04-14	878035	Video Game Arcades/Establishments
5	7995	8590128346463420	439.52	2003-04-11	444413	Betting (including Lottery Tickets, Casino Gaming
6	7995	8590128347406193	749.14	2003-04-14	778088	Betting (including Lottery Tickets, Casino Gaming

O resultado obtido demonstra que 7 transações ocorridas em abril foram transações inaceitáveis, melhor explicado, na tabela trans_abril há 7 transações com códigos que fazem parte da tabela cod_naoAceitos.

Identificando os funcionários:

Agora que você confirmou que algumas transações tem códigos inaceitáveis, você determina quais os funcionários fizeram essas compras. Você também pode identificar o departamento ao qual esses funcionários pertencem. Você pode obter informações de várias tabelas e combiná-lo em uma única visão. O objetivo é relatar:

■ Nome do empregado

■ número do cartão de crédito

■ Departamento

■ Quantidade de transação

■ Quantidade de transação

■ Data de transação

■ código de transação

■ Descrição do código de transação

Identificar o departamento dos funcionários:

Seu primeiro passo é associar cada funcionário ao seu departamento.

■ A tabela Company_Dept contém códigos e nomes de departamento, como 800000 para Consultoria.

■ A tabela Empregados contém o campo EmpNo, que começa com o primeiro dígito do código do departamento.

Ao comparar o primeiro dígito do código do empregado ao primeiro dígito do código do departamento da empresa, você pode identificar o departamento ao qual o empregado pertence.

Para fazer essa comparação, você cria um novo campo em cada tabela que recebe o primeiro dígito dos valores Dept_Code. Em seguida, você compara esse dígito com o primeiro dígito do EmpNo na tabela Empregados.

In [48]:

```
dep_cia.head()
```

Out[48]:

	Dept_Name	Dept_Code
0	Administrative	000000
1	Product Development	100000
2	Manufacturing	200000
3	Technical Support	300000
4	Sales	400000

In [49]:

```
dep_cia['Novo_Cod'] = [num[0] for num in dep_cia.Dept_Code]
dep_cia.head()
```

Out[49]:

	Dept_Name	Dept_Code	Novo_Cod
0	Administrative	000000	0
1	Product Development	100000	1
2	Manufacturing	200000	2
3	Technical Support	300000	3
4	Sales	400000	4

Faremos o mesmo com a coluna EmpNo do banco de dados employees

In [50]:

```
employees.head()
```

Out[50]:

	First_Name	Last_Name	CardNum	EmpNo	HireDate	Salary	Bonus_2002
0	Leila	Remlawi	8590122497663800	8	1997-12-28	52750	1,405.40
1	Vladimir	Alexov	8590122281964010	60	1997-10-05	41250	4,557.43
2	Matthew	Lee	8590120784984560	100	1999-03-31	38250	651.19
3	Alex	Williams	8590124253621740	104	2001-08-12	40175	7,460.02
4	Narinder	Singh	8590125999743360	146	1999-09-09	32250	6,990.75

In [51]:

```
employees['Novo_Cod'] = [num[0] for num in employees.EmpNo]
employees.sort_values('Novo_Cod').head()
```

Out[51]:

	First_Name	Last_Name	CardNum	EmpNo	HireDate	Salary	Bonus_2002	Novo_Cod
8	Jeanette	Wallace	8590128676326310	180	1995-05-11	46500	952.81	1
28	John	Gilbert	8590125323000750	100093	1998-09-08	35900	4,783.03	1
29	Janet	Gilley	8590128098022580	100125	1999-07-02	29950	6,325.61	1
30	Lars	Andersson	8590126009071520	100130	1995-02-18	45370	6,000.45	1
31	Olga	Roman	8590128742834790	100188	1996-04-30	63840	2,596.77	1

In [52]:

```
# Juntando as tabelas employees e dep_cia pela coluna "Novo_Cod" encontrada em ambas as tabelas, para que possamos trazer
# para a tabela "employees" o nome do Departamento.
empr_dep = pd.merge(dep_cia,employees,
                    on = 'Novo_Cod', how = 'inner')
empr_dep['Bonus_2002'] = empr_dep['Bonus_2002'].apply(lambda x: x.replace('$', '').replace(',', '').astype('float'))
empr_dep.head()
```

Out[52]:

	Dept_Name	Dept_Code	Novo_Cod	First_Name	Last_Name	CardNum	EmpNo	HireDate	Salary	Bonus_2002
0	Product Development	100000	1	Matthew	Lee	8590120784984560	100	1999-03-31	38250	651.19
1	Product Development	100000	1	Alex	Williams	8590124253621740	104	2001-08-12	40175	7460.02
2	Product Development	100000	1	Narinder	Singh	8590125999743360	146	1999-09-09	32250	6990.75
3	Product Development	100000	1	Albert	Schmidt	8590120716753180	157	2002-09-26	36170	836.98
4	Product Development	100000	1	Mohan	Parhar	8590128947747850	161	2000-08-10	69750	4455.37

Agora estamos prontos para acrescentar o nome, sobrenome e departamento a tabela que contém as transações não aceitas.

In [53]:

```
trans_abril_naoAceitos.head(3)
```

Out[53]:

	CODES	CARDNUM	AMOUNT	TRANS_DATE	CUSTNO	DESCRIPTION
0	4411	8590126009071523	314.33	2003-04-14	284354	Cruise Lines
1	4411	8590126069373535	226.08	2003-04-14	444413	Cruise Lines
2	5094	8590126961897004	745.88	2003-04-07	444413	Precious Stones and Metals, Watches and Jewelry

In [54]:

```
trans_abril_naoAceitos = trans_abril_naoAceitos.rename(columns={'CARDNUM':'CardNum'}) # Deixando o nome igual ao da
# tabela empr_dep
```

In [55]:

```
trans_abril_naoAceitos['CardNum'] = trans_abril_naoAceitos['CardNum'].astype('object')
trans_abril_naoAceitos['CODES'] = trans_abril_naoAceitos['CODES'].astype('object')
trans_abril_naoAceitos['CUSTNO'] = trans_abril_naoAceitos['CUSTNO'].astype('object')
trans_abril_naoAceitos.head()
```

Out[55]:

	CODES	CardNum	AMOUNT	TRANS_DATE	CUSTNO	DESCRIPTION
0	4411	8590126009071523	314.33	2003-04-14	284354	Cruise Lines

1	CODES	CardNum	AMOUNT	TRANS_DATE	CUSTNO	DESCRIPTION
2	5094	8590126961897004	745.88	2003-04-07	444413	Precious Stones and Metals, Watches and Jewelry
3	5094	8590127015016937	457.70	2003-04-09	962353	Precious Stones and Metals, Watches and Jewelry
4	7994	8590128345724706	642.83	2003-04-14	878035	Video Game Arcades/Establishments

In [56]:

```
trans_abril_naoAceitos_consolidado = pd.merge(trans_abril_naoAceitos, empr_dep,
                                              on = 'CardNum', how = 'left');
```

In [58]:

```
trans_abril_naoAceitos_consolidado
```

Out[58]:

	CardNum	CODES	AMOUNT	TRANS_DATE	CUSTNO	DESCRIPTION	Dept_Name	First_Name	Last_Name
0	8590126009071520	4411	314.33	14/04/2003	284354	Cruise Lines	Product Development	Lars	Andersson
1	8590126069373530	4411	226.08	14/04/2003	444413	Cruise Lines	Technical Support	Sarah	Pollard
2	8590126961897000	5094	745.88	07/04/2003	444413	Precious Stones and Metals, Watches and Jewelry	Consulting	Tina	Fisher
3	8590127015016930	5094	457.70	09/04/2003	962353	Precious Stones and Metals, Watches and Jewelry	Consulting	Reba	John
4	8590128345724700	7994	642.83	14/04/2003	878035	Video Game Arcades/Establishments	Consulting	Dian	Hemminger
5	8590128346463420	7995	439.52	11/04/2003	444413	Betting (including Lottery Tickets, Casino Gaming)	Marketing	Sean	Brynnlfsson
6	8590128347406190	7995	749.14	14/04/2003	778088	Betting (including Lottery Tickets, Casino Gaming)	Technical Support	Valerie	Cooper

Agora você pode ver a lista de funcionários cujos cartões de crédito foram usados para cobrar transações inaceitáveis. Naturalmente, você não pode assumir imediatamente que os funcionários listados são responsáveis pelas transações inaceitáveis porque é possível que alguém tenha usado seus cartões de crédito. Para isso necessitaríamos de posteriores análises.

Idéias para análises posteriores: Nesta demonstração, você trabalhou com vários exemplos de como realizar análises usando a linguagem Python. Quais outras análises você poderia realizar com os dados? Considere estas questões. É provável que você tenha várias ideias aplicáveis ao seu próprio trabalho.

Algumas análises que podem ser feitas:

- As transações têm códigos de categoria do comerciante inaceitáveis?
- Os códigos de categoria do comerciante usados correspondem às responsabilidades do portador do cartão?
- O titular do cartão realiza transações repetidas em intervalos regulares com o mesmo comerciante?
- O titular do cartão tem várias transações em um breve período de tempo (digamos, 72 horas) com o mesmo comerciante?
- O titular do cartão tem transações que ocorrem nos fins de semana ou feriados?
- Um número incomum de transações arredondou valores, como \$ 1000, \$ 200 ou \$ 450?
- Existem créditos cobrados de volta a um número de cartão e quais são as explicações para esses créditos?
- O portador do cartão chega perto ou excede seu limite de crédito?
- Houve um aumento repentino nas compras do portador do cartão?
- O titular do cartão é responsável por cobrar as despesas de outras pessoas?
- Quais cinco ou dez funcionários gastaram mais? Para quais clientes eles cobraram suas despesas? Existe alguma relação entre gastar muito e cobrar transações inaceitáveis?
- Determinar o gasto total por departamento mensalmente. Use o campo DeptCode na tabela Empregados.
- Se um empregado cobra tarifa aérea, que outras despesas relacionadas a viagens ele ou ela cobraram? Por exemplo, se um funcionário cobra tarifa aérea, é provável que ele cobra um quarto de hotel e o transporte. É este o caso? Quais são algumas explicações possíveis para os padrões

taínia aérea, e provável que ele cobra um quarto de hotel e o transporte. E este o caso? Quais são algumas explicações possíveis para os padrões que aparecem?

■ Se os códigos para fornecedores e padarias fossem inaceitáveis, quanto dinheiro teria sido economizado?

Bom, espero que esse trabalho tenha colaborado para demonstrar a importância de saber analisar e manipular dados. Até o próximo!!