

# 6 month expense analysis for 2019

Zachary Yeap

Last updated: 10/01/19

This project uses expense data from January to June of 2019. Each expense has an id, month, amount, and category. Expenses are divided into 6 categories: clothing, entertainment, food, technology, transportation, and miscellaneous.

With this project, I aim to perform basic data analysis on my expenses within the 6 month timeframe. Data analysis will be performed on each month to find total amounts and the proportion of spending for each category.

The raw Excel spreadsheet data can be found at: [Expensedata2019.xlsx](#)

```
In [267]: #importing libraries  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt
```

```
In [268]: x1 = pd.ExcelFile("Expensedata2019.xlsx")
ExpenseData = x1.parse("Expenses")
ExpenseData["month"] = pd.to_datetime(ExpenseData.month, format = "%b", errors = "coerce").dt.month

MonthlyExpense = ExpenseData[["month", "amount"]]
MonthlyExpenseSum = MonthlyExpense.groupby(by = "month").sum().sort_values(by = "month")
formatted_month = MonthlyExpenseSum.reset_index()

print("---Total amount spent per month---\n")
print(formatted_month)
print("\n")

TotalExpenseData = ExpenseData[["amount"]].sum()
print("Total amount spent from January to June: ")
print(TotalExpenseData)
print("\n")

Amount_series = MonthlyExpenseSum[["amount"]]
AvgExpense = Amount_series.mean()
print("Average amount spent over 6 months: ")
print(AvgExpense)
```

---Total amount spent per month---

	month	amount
0	1	259.10
1	2	175.01
2	3	164.44
3	4	266.79
4	5	805.69
5	6	285.76

Total amount spent from January to June:  
amount 1956.79  
dtype: float64

Average amount spent over 6 months:  
amount 326.131667  
dtype: float64

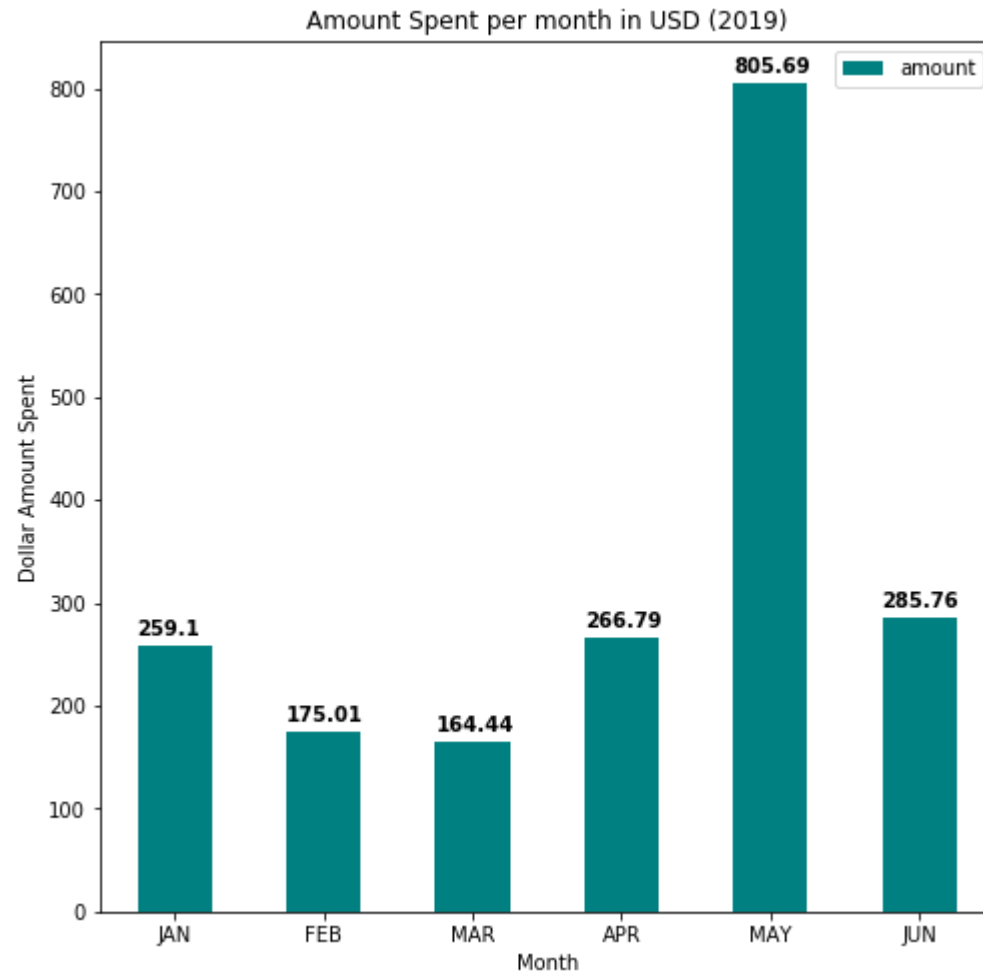
```
In [266]: values = [259.10, 175.01, 164.44, 266.79, 805.69, 285.76]

MonthlyExpChart = MonthlyExpense.groupby("month").sum().plot(kind = "bar", color = "teal")

MonthlyExpChart.set_title("Amount Spent per month in USD (2019)")
MonthlyExpChart.set_ylabel("Dollar Amount Spent")
MonthlyExpChart.set_xlabel("Month")
MonthlyExpChart.set_xticklabels(("JAN", "FEB", "MAR", "APR", "MAY", "JUN"))

for i, v in enumerate(values):
    MonthlyExpChart.text(i - .25, v + 10, str(v), fontweight = "bold")

plt.rcParams["figure.figsize"] = [8,8]
plt.xticks(rotation=0)
plt.show()
```

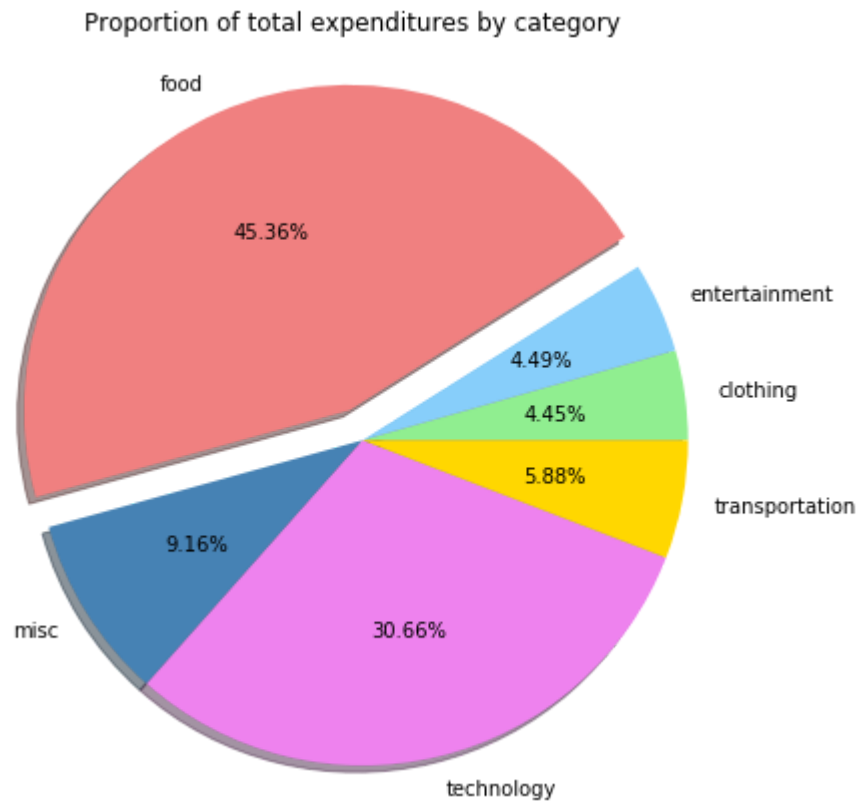


```
In [254]: ExpenseData.reset_index()
df = ExpenseData[["category", "amount"]]
Exp_cat = df.groupby(by = "category").sum().reset_index()
print("---Total expense amount per category---\n")
print(Exp_cat)
print("\n")

plt.figure(figsize = (7,7))
plt.title("Proportion of total expenditures by category")
colors = ["lightgreen", "lightskyblue", "lightcoral", "steelblue", "violet", "gold"]
explode = (0, 0, .1, 0, 0, 0)
plt.pie(Exp_cat.amount, explode = explode, labels = Exp_cat.category, autopct = "%.2f%", colors = colors, shadow = True)
plt.axis('equal')
plt.show()
```

---Total expense amount per category---

	category	amount
0	clothing	87.17
1	entertainment	87.87
2	food	887.55
3	misc	179.20
4	technology	600.00
5	transportation	115.00



## Analysis for May 2019

Expenses were the highest for May, thus an expenditure breakdown will be performed for this month.

```
In [264]: ExpenseData.reset_index()
df_2 = ExpenseData[["month", "category", "amount"]]
MayExp = df_2.loc[df_2["month"] == 5]
MayExp_cat = MayExp[["category", "amount"]].groupby(by = "category").sum().reset_index()
print("---Expenditure breakdown for May 2019---\n")
print(MayExp_cat)
print("\n")

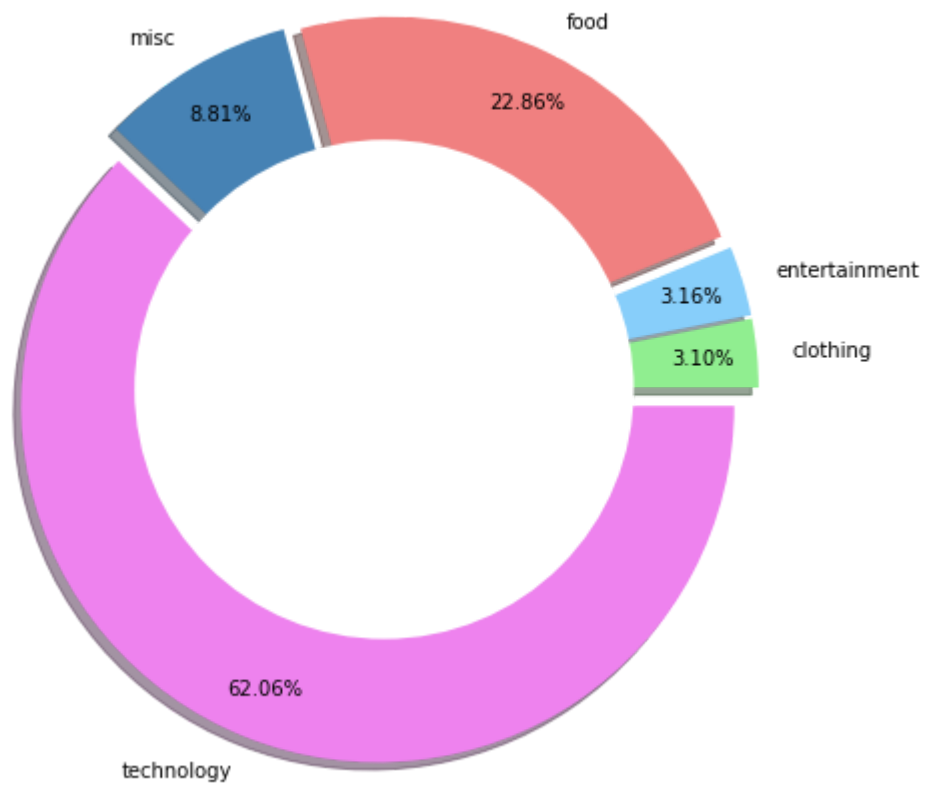
plt.figure(figsize = (7,7))
plt.title("Proportion of total expenditures by category for May 2019")
explode = (.05,.05,.05,.05,.05)
plt.pie(MayExp_cat.amount, explode = explode, labels = MayExp_cat.category, autopct = "%.2f%%",
        pctdistance = 0.85, colors = colors, shadow = True)
plt.axis('equal')
plt.tight_layout()

donut = plt.Circle((0,0), 0.70, fc = "white")
ref = plt.gcf()
ref.gca().add_artist(donut)
plt.show()
```

---Expenditure breakdown for May 2019---

	category	amount
0	clothing	25.00
1	entertainment	25.50
2	food	184.19
3	misc	71.00
4	technology	500.00

Proportion of total expenditures by category for May 2019





## Analysis of food expenditures

Food is the largest expense over the 6 months, so I wanted to track the amount spent on food per month.

```
In [304]: ExpenseData.reset_index()

#create dataframe with month, category, and amount columns
category_df = ExpenseData[["month", "category", "amount"]]

#find values where category is only food
food_df = category_df.loc[category_df["category"] == "food"]

#create new dataframe with month and amount for only food expenses
foodsum_df = food_df[["month", "amount"]].groupby(by = "month").sum().reset_index()
print(foodsum_df)

labels = ["fill", "JAN", "FEB", "MAR", "APR", "MAY", "JUN"]
food_lplot = sns.lineplot(x = "month", y = "amount", data = foodsum_df)
food_lplot.set_title("Food expenditure amount from January to June 2019")
food_lplot.set(xlabel = "Month", ylabel = "Dollar Amount Spent")
food_lplot.set(xticklabels = labels)
plt.show()
```

	month	amount
0	1	164.10
1	2	99.44
2	3	95.44
3	4	185.79
4	5	184.19
5	6	158.59

