Energy Use with Smart Meters

A Study of Electricity Consumption in England, Wales and Scotland

**By: Sau Kha**

**12/14/2017**

# Executive Summary

The UK government requires energy providers to provide smart meters in each home in England, Wales and Scotland. The goal is for each home to have a smart meter by 2020, which means that eventually, more than 26 million homes will be providing data for the energy providers.

While the huge project is on the way, a pilot project was conducted from November 2011 and February 2014 to investigate better energy use. This report documents the research questions, methodology, findings and conclusions from the pilot study. Lessons learned, recommendations and scope of further study are also covered.

The research questions are:

1. According to the Acorn classification, which category of household is the biggest user of electricity for the time frame of the study?
2. Within the category, which type of household is the biggest user of electricity?
3. Which are the peak usage months?

Fifteen blocks of huge data from smart meters of 5,567 London households were explored, analyzed and visualized. Python programming tools, mainly the Pandas and Matplotlib modules, were used.

Appendix A presents all the Python scripts used in this study. Appendix B displays charts generated from the dataset.

# Table of Contents

[Executive Summary i](#_Toc501101362)

[Introduction 1](#_Toc501101363)

[Research Questions 1](#_Toc501101364)

[Reference 1](#_Toc501101365)

[Methodology 2](#_Toc501101366)

[Findings 4](#_Toc501101367)

[Re-Grouping of Acorn Types in Broader Categories 4](#_Toc501101368)

[Answer to Research Question 1 5](#_Toc501101369)

[Answer to Research Question 2 6](#_Toc501101370)

[Answer to Research Question 3 7](#_Toc501101371)

[Conclusion 7](#_Toc501101372)

[APPENDIX A: Python Scripts 8](#_Toc501101373)

[APPENDIX B: Charts 18](#_Toc501101374)

[Figures 1 - 16: Total Electricity Usage by Acorn Category Per Data Block 19](#_Toc501101375)

[Figures 17 – 32: Total Electricity Usage by Acorn Category by Month Per Data Block 27](#_Toc501101376)

[Figures 33 – 48: Total Electricity Usage by Acorn Type Per Data Block 35](#_Toc501101377)

List of Charts and Tables

[Table 1. Distribution of Smart Meter Users in Acorn Category 4](#_Toc501101406)

[Chart A. Total Electricity Usage by Acorn Category 5](#_Toc501101407)

[Table 2. Overall Usage Distribution by Acorn Category 5](#_Toc501101408)

[Chart B. Total Electricity Usage by Acorn Type 6](#_Toc501101409)

[Table 3. Overall Usage Distribution by Acorn Type 6](#_Toc501101410)

List of Figures

Figure 1: Total Electricity Usage by Acorn Category – Block\_0-a 19

Figure 2: Total Electricity Usage by Acorn Category – Block\_0-b 19

Figure 3: Total Electricity Usage by Acorn Category – Block\_1 20

Figure 4: Total Electricity Usage by Acorn Category – Block\_2 20

Figure 5: Total Electricity Usage by Acorn Category – Block\_3 21

Figure 6: Total Electricity Usage by Acorn Category – Block\_4 21

Figure 7: Total Electricity Usage by Acorn Category – Block\_5 22

Figure 8: Total Electricity Usage by Acorn Category – Block\_6 22

Figure 9: Total Electricity Usage by Acorn Category – Block\_7 23

Figure 10: Total Electricity Usage by Acorn Category – Block\_8 23

Figure 11: Total Electricity Usage by Acorn Category – Block\_9 24

Figure 12: Total Electricity Usage by Acorn Category – Block\_10 24

Figure 13: Total Electricity Usage by Acorn Category – Block\_11 25

Figure 14: Total Electricity Usage by Acorn Category – Block\_12 25

Figure 15: Total Electricity Usage by Acorn Category – Block\_13 26

Figure 16: Total Electricity Usage by Acorn Category – Block\_14 26

Figure 17: Total Electricity Usage by Acorn Category by Month – Block\_0-a 27

Figure 18: Total Electricity Usage by Acorn Category by Month – Block\_0-b 27

Figure 19: Total Electricity Usage by Acorn Category by Month – Block\_1 28

Figure 20: Total Electricity Usage by Acorn Category by Month – Block\_2 28

Figure 21: Total Electricity Usage by Acorn Category by Month – Block\_3 29

Figure 22: Total Electricity Usage by Acorn Category by Month – Block\_4 29

Figure 23: Total Electricity Usage by Acorn Category by Month – Block\_5 30

Figure 24: Total Electricity Usage by Acorn Category by Month – Block\_6 30

Figure 25: Total Electricity Usage by Acorn Category by Month – Block\_7 31

Figure 26: Total Electricity Usage by Acorn Category by Month – Block\_8 31

Figure 27: Total Electricity Usage by Acorn Category by Month – Block\_9 32

Figure 28: Total Electricity Usage by Acorn Category by Month – Block\_10 32

Figure 29: Total Electricity Usage by Acorn Category by Month – Block\_11 33

Figure 30: Total Electricity Usage by Acorn Category by Month – Block\_12 33

Figure 31: Total Electricity Usage by Acorn Category by Month – Block\_13 34

Figure 32: Total Electricity Usage by Acorn Category by Month – Block\_14 34

Figure 33: Total Electricity Usage by Acorn Type – Block\_0-a 35

Figure 34: Total Electricity Usage by Acorn Type – Block\_0-b 35

Figure 35: Total Electricity Usage by Acorn Type – Block\_1 36

Figure 36: Total Electricity Usage by Acorn Type – Block\_2 36

Figure 37: Total Electricity Usage by Acorn Type – Block\_3 37

Figure 38: Total Electricity Usage by Acorn Type – Block\_4 37

Figure 39: Total Electricity Usage by Acorn Type – Block\_5 38

Figure 40: Total Electricity Usage by Acorn Type – Block\_6 38

Figure 41: Total Electricity Usage by Acorn Type – Block\_7 39

Figure 42: Total Electricity Usage by Acorn Type – Block\_8 39

Figure 43: Total Electricity Usage by Acorn Type – Block\_9 40

Figure 44: Total Electricity Usage by Acorn Type – Block\_10 40

Figure 45: Total Electricity Usage by Acorn Type – Block\_11 41

Figure 46: Total Electricity Usage by Acorn Type – Block\_12 41

Figure 47: Total Electricity Usage by Acorn Type – Block\_13 42

Figure 48: Total Electricity Usage by Acorn Type – Block\_14 42

# Introduction

To study the energy consumption of the country, energy providers in the UK introduce smart meters in the homes in England, Wales and Scotland. The plan is that by 2020, more than 26 million homes will be equipped by smart meters that log electricity usage data continuously 24 hours, 7 days a week. Collected data will be analyzed to formulate better energy usage strategy.

This document reports findings from a sample of 5,567 London households that took part in a pilot project between November 2011 and February 2014.

Smart meter data from participant households and their household classification based on an ACORN approach [Reference 1 and 2] are analyzed. Findings in relation to amount and pattern of usage may be used to strategically plan future energy development.

## Research Questions

Research questions for this study include:

1. According to the Acorn classification, which category of household is the biggest user of electricity for the time frame of the study?
2. Within the category, which type of household is the biggest user of electricity?
3. Which are the peak usage months?

After exploring the collected data, plans were to acquire additional data, such as weather and life styles of household to explore possible explanations for the observations.

Intended audience of this study include government policy makers, energy providers, energy conservation groups and electricity consumers. Findings from the study may be used to make strategic energy development plan for different areas of the country and consumer-targeted promotions of energy savers consumer plans, energy efficient cars and appliances, energy generators, solar panels, and so on. This type of study should be conducted in a continuous bases, so that strategic plans may be adjusted as the situation changes.

# Reference

1. “The Acorn User Guide, The Consumer Classification”, CACI Ltd, 2014.
2. “Acorn Technical Guide”, CACI Ltd, 3/2/2017.

# Methodology

The methodology for this study basically is first to explore, clean up, analyze, visualize the data and then draw conclusions and attempt to find answers to the research questions. The following steps were taken to reach the findings and conclusions:

1. Unzip the data files and write Python scripts to explore and clean up data. To progressively approach the study, data files were written over with cleaned up file versions. These include all 15 blocks of smart meter data and the information household filed.
   1. Data Clean Up: Any null, blank and erroneous record are dropped. This include user data with incomplete Acorn Types, such as ACORN- and ACORN-U, that do not relate to any Acorn Category.

Lesson learned: This was done because most of the time they cause error when a Python function or method is attempted.

* 1. Information Extraction and Exploration: The timestamp column from the text smart reader files were converted to datetime datatype. As each data file was explored and cleaned up, I used Pandas function (pd.datetime) to extract Year, Month and Hour information from the datetime column. Year, Month and Hour information were saved as added columns in the csv file for later data analysis.

Lesson Learned: The additional columns were not necessary, as the same information could be obtained during data analysis phase. This also have made the source text files bigger and cost longer or memory problem during analysis. For example, the first zip file of 15 blocks of smart meter data, block\_0.csv.gz, has the largest file size. After extracting the and adding extra columns, the resulting text file when read as a Pandas DataFrame, that file was the one of the few that consistently caused my computer to freeze. The IPython kernel, the Python execution backend for Jupyter notebook frequently died or automatically restarted.

* 1. Data Merging and Analysis: Household information of participants were merged with the smart meter data, so that data could be analyzed per Acorn Type and Category. Lesson Learned:
     1. While it was possible read and analyze data with Pandas, it would be much better if a relational database was created to import data from each block as a table in the database. Together with the household information imported as a separate table, queries could be done to pull data from different tables for better data analysis. SQLite3 is a very good tool for this type of analysis.
     2. With memory limitation, I considered using the pickle module in Python. The idea was to convert a slice of information, such as the total electricity usage for a certain time frame for a certain Acorn type or category, from a block of smart meter data into Pandas DataFrame and then to save it as pickle file. After the same slice of information was pulled from each block and saved as a pickle file, all pickle could be concatenated and visualized.

The [pickle](https://docs.python.org/3/library/pickle.html" \l "module-pickle) module implements binary protocols for serializing and de-serializing a Python object structure. “Pickling” is the process whereby a Python object hierarchy is converted into a byte stream, and “unpickling” is the inverse operation, whereby a byte stream (from a [binary file](https://docs.python.org/3/glossary.html" \l "term-binary-file) or [bytes-like object](https://docs.python.org/3/glossary.html" \l "term-bytes-like-object)) is converted back into an object hierarchy. Pickling (and unpickling) is alternatively known as “serialization”, “marshalling,” or “flattening”; however, to avoid confusion, the terms used here are “pickling” and “unpickling”. *Source: Python Standard Library 12.1 Python Object Serialization”* [***https://docs.python.org/3/library/pickle.html***](https://docs.python.org/3/library/pickle.html)

* + 1. With memory problem encountered early in the project, I decided to just analyze each block of smart meter data separately. If each block of data was an unbiased sample of the population, the analysis result from each block should be comparable to each other. Otherwise, further analysis would be needed to obtain overall results from the complete dataset. In the last minutes as I was writing this report, I was able to use Pandas “groupby” mechanism to systematically aggregate summary data and save them in two csv files in Python. That was completed by writing and running scripts directly on Jupyter Notebook. The next set of groupby scripts were run by revising what were on an untitled Jupyter Notebook. This could have been done probably in a well-planned and efficient programming way.
  1. Data Visualization: Charts were created from each block of data and compared among each other using the Matplotlib module.
  2. Draw Conclusion: Find answers to the research questions of this study and draw conclusions.

# Findings

## Re-Grouping of Acorn Types in Broader Categories

Table 1 presents the distribution of smart meter users by Acorn Category and Type. The cleaned-up dataset consists of users from only three Acorn categories: Affluent (40%), Comfortable (27%) and Adversity (33%). Eight (highlighted) of the 15 blocks of data consist 40% or more users of the Affluent Category.

A closer look at the household information file reviewed that the grouping of Acorn Types into categories in this dataset differs from the grouping system indicated in the Acorn User Guide (Reference 1). Types A through C (Affluent Achievers) were grouped with Types D and E (Rising Prosperity) as the Affluent, while Types K through N (Financially Stretched) were grouped with Types O through Q (Urban Adversity) as Adversity.

When comparing the energy usage between categories, readers should bear this regrouping in mind. Refer to the Acorn User Guide (Reference 1) for further information of each Acorn Type and Category.

Table 1. Distribution of Smart Meter Users in Acorn Category

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Acorn Categories / Types | | |
| file | Total | Affluent  (Types A – E) | Comfortable  (Types F -J) | Adversity  (Types K – Q) |
| block\_0 | 388 | 45% | 26% | 28% |
| block\_1 | 390 | 36% | 24% | 40% |
| block\_2 | 388 | 34% | 27% | 39% |
| block\_3 | 388 | 37% | 27% | 36% |
| block\_4 | 386 | 38% | 24% | 38% |
| block\_5 | 387 | 42% | 24% | 34% |
| block\_6 | 384 | 30% | 34% | 36% |
| block\_7 | 386 | 22% | 41% | 38% |
| block\_8 | 384 | 40% | 28% | 33% |
| block\_9 | 383 | 43% | 23% | 34% |
| block\_10 | 388 | 56% | 22% | 22% |
| block\_11 | 387 | 45% | 23% | 32% |
| block\_12 | 386 | 47% | 25% | 27% |
| block\_13 | 385 | 38% | 35% | 27% |
| block\_14 | 105 | 51% | 28% | 21% |
| Overall | 5515 | 40% | 27% | 33% |

## Answer to Research Question 1

The first question is: according to the Acorn classification, which category of households is the biggest users of electricity for the time frame of the study?

The Affluent Category is the biggest group of users. Chart A and Table 2 below present the total electricity usage by Acorn Category. The Affluent users used 46%, while the other two groups tied with 27% usage of all electricity consumption during the time of study. Figures 1 through 16 in Appendix B present the Total Electricity Usage by Acorn Category. The charts indicate that in all data blocks, the Affluent Category consistently resulted in the highest usage, except for Blocks 6 and 7 (see Figures 8 and 9 in Appendix B. This is because there were less Affluent users in these blocks when compared with other two categories. Additionally, Affluent users in these two blocks were the least of all 15 blocks.

Chart A. Total Electricity Usage by Acorn Category

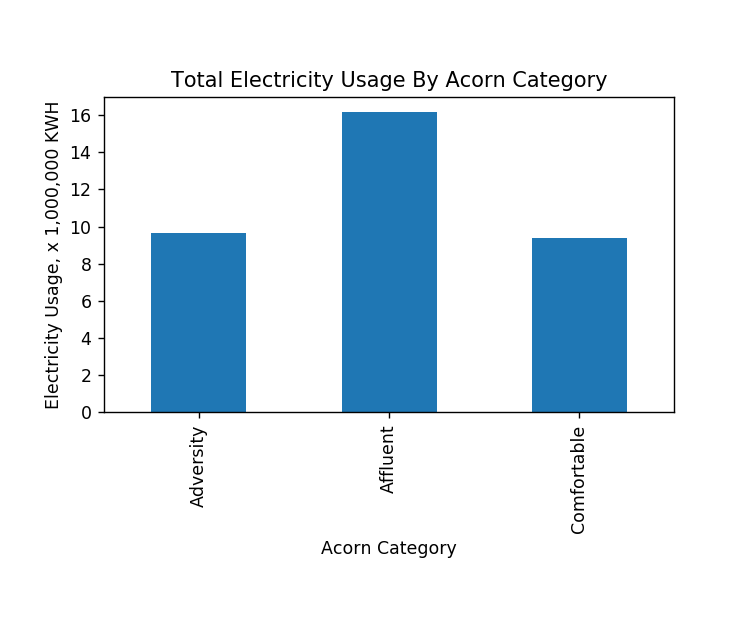


Table 2. Overall Usage Distribution by Acorn Category

|  |  |  |  |
| --- | --- | --- | --- |
| Acorn Category | Acorn Type | Total Usage (KWH) | Percent |
| Affluent | Type A - E | 16158191 | 46% |
| Adversity | Type K - Q | 9654154 | 27% |
| Comfortable | Type F -J | 9374789 | 27% |

## Answer to Research Question 2

The second research question is: Within the category, which type of household is the biggest user of electricity?

Acorn Type E (Career Climbers) is the biggest group of users. Chart B and Table 3 below present the total electricity usage by Acorn Type. Among the Acorn types that belong to the Affluent Category, Type E (Career Climbers) are the biggest electricity consumers and used 30% of the total usage for the time frame of the study. Additionally, as shown in Figures 33 through 48 in Appendix B, Acorn Type E users are also big spenders of electricity in all data blocks, except in Blocks 6 and 7, which could possibly be explained by the same reason in the section above.

Chart B. Total Electricity Usage by Acorn Type

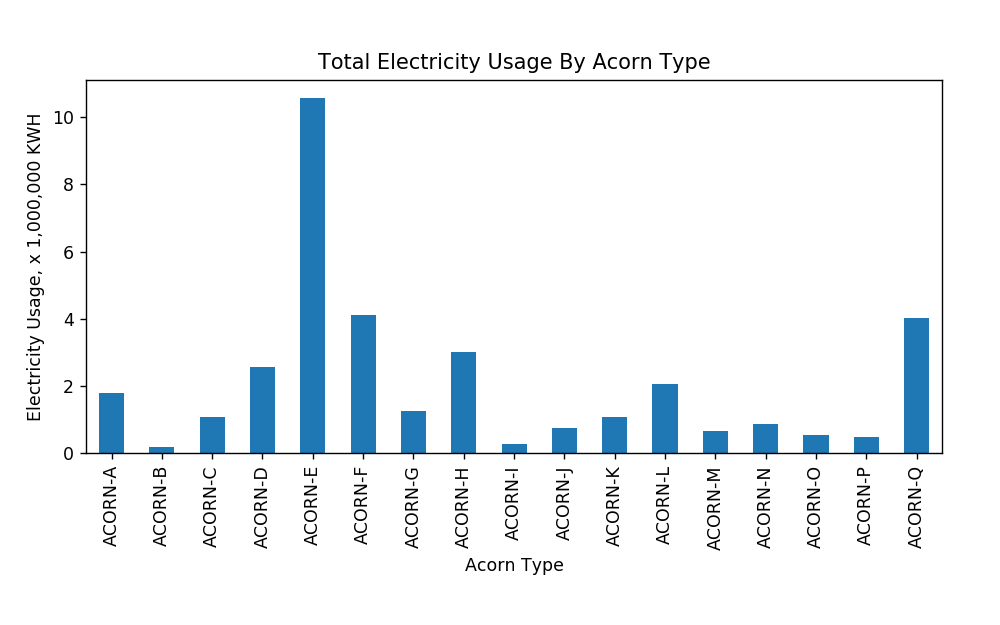


Table 3. Overall Usage Distribution by Acorn Type

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Acorn **Re-Grouped** Categories | Acorn Types | | Total Usage (KWH) | | | Percent |
| Affluent | ACORN-A | Lavish Lifestyles | | 1774523 | 5.0% | |
| Affluent | ACORN-B | Executive Wealth | | 172171 | 0.5% | |
| Affluent | ACORN-C | Mature Money | | 1059742 | 3.0% | |
| Affluent | ACORN-D | City Sophisticates | | 2565846 | 7.3% | |
| Affluent | ACORN-E | Career Climbers | | 10585909 | 30.1% | |
| Comfortable | ACORN-F | Countryside Communities | | 4110611 | 11.7% | |
| Comfortable | ACORN-G | Successful Suburbs | | 1257513 | 3.6% | |
| Comfortable | ACORN-H | Steady Neighbourhoods | | 2992531 | 8.5% | |
| Comfortable | ACORN-I | Comfortable Seniors | | 278233 | 0.8% | |
| Comfortable | ACORN-J | Starting Out | | 735901 | 2.1% | |
| Adversity | ACORN-K | Student Life | | 1066916 | 3.0% | |
| Adversity | ACORN-L | Modest Means | | 2045429 | 5.8% | |
| Adversity | ACORN-M | Striving Families | | 661599 | 1.9% | |
| Adversity | ACORN-N | Poorer Pensioners | | 847986 | 2.4% | |
| Adversity | ACORN-O | Young Hardship | | 522308 | 1.5% | |
| Adversity | ACORN-P | Struggling Estates | | 477079 | 1.4% | |
| Adversity | ACORN-Q | Difficult Circumstances | | 4032837 | 11.5% | |

## Answer to Research Question 3

The third question is: Which are the peak usage months?

Figures 17 through 32 in Appendix B display charts showing the total electricity usage by Acorn Category by month per data block. January and December are the two months of peak usage in each data block and each Acorn Category. Usage gradually declines from January towards warmer months in the Summer and then gradually climbs in mid-year and peaks in December.

To substantiate and relate the electricity usage to weather condition, it is recommended to acquire temperature data from the time of study for the country. API is available from Darksky.net for this purpose. Users may open accounts for free and make 1,000 API requests to the Service per day free of charge. The cost for API calls past 1,000 per day is also minimal at $0.0001 USD per API request. API requests to the Service return data in JSON format.

# Conclusion

In short, analysis of the dataset provides answers to all three research questions. Affluent consumers are the biggest users of electricity. And within the Affluent category, Acorn Type E, Career Climbers, use the most energy. Additional information concerning characteristics and lifestyles of Career Climbers may be obtained from the Acorn User Guide (Reference 1).

People’s career opportunity, family situation, financial obligations and living area and housing styles could very well explain the high energy usage of this group of people. Besides, participants of the Smart Meter pilot study did not acquire the meter at the same time. So, comparing the total usage from among Acorn categories or types may not be comparing apples to apples. Participants who started in the winter months would have logged more usage in the meter than those who started in the summer months. Likewise, those who participated in the early phase of the pilot study will have a greater total usage. Use of more sophisticated programming tools will be able to assist with more precise data analysis. The total number of individuals in Acorn categories and types will certainly affect the total usage for the group. It is better to the calculate the average usage by dividing the total usage by the number of participating household. The number of family members in the household, the region and the type of housing, and so on, may have an impact on energy usage. However, this type of information is not available for the study.

As far as the peak usage in relation the weather in certain months of the year, it was pretty much predicted. Future work to continue this study includes investigating the use of API to collect weather data.

The major weakness of the study is the lack of proficiency in Python packages, such as sqlite3 or SQLAlchemy. These tools will allow queries to be run to gather relational information from different tables. In the case of the current study, information from different data blocks, may be assembled according to Acorn Category, Type, year, month or time of day and analyzed.

# APPENDIX A: Python Scripts

Project3\_improved\_modules.py

import pandas as pd

import numpy as np

import matplotlib

import datetime

import pickle

import sqlite3

from matplotlib import pyplot as plt

# %matplotlib inline # this code is only to plot inline in Jupyter Notebook

#read, clean and format data

def clean\_format\_csv(filename):

df=pd.read\_csv(filename, na\_values=[' ','Null','null','N/A','na','',float('inf'), float('-inf'), np.inf, -np.inf])

print (filename, df.shape)

df.drop\_duplicates(inplace=True)

df.dropna(inplace=True)

df.DateTime=pd.to\_datetime(df.DateTime) #convert DateTime column from str to DateTime dtype

df['Year']=df.DateTime.dt.year #add 'Year' column based on Year valud in DateTime col

df['Month']=df.DateTime.dt.month #add 'Month' column based on Month value in DateTime col

df['Day\_of\_Week (Mon=0 .. Sun=6)']=df.DateTime.dt.dayofweek #add 'Date\_of\_Week [Mon=0 .. Sun=6]' column based on dayofweek value in DateTime col

df['Hour']=df.DateTime.dt.hour #add 'Hour' column based on Hour valud in DateTime col

df.to\_csv(filename, sep=',') # overwrite data file with cleaned up version with added columns

print (filename, df.shape)

# merge dfs

def merge\_blockdata\_and\_type(blockfilename,typefilename):

df\_merged=pd.DataFrame()

df1=pd.read\_csv(blockfilename)

df2=pd.read\_csv(typefilename)

print (df1.head())

print (df2.head())

df\_merged=pd.merge(df1,df2,how='left',on='LCLid')

print (df\_merged.tail())

print type(df\_merged)

print df\_merged.dtypes

return df\_merged # failed to return as pd.DataFrame type #why?

# analyze and visualize data one block at a time and merge dfs

# enter one filename at a time and run this scripts

def analyze\_visualize (filename):

block=filename

df1=pd.read\_csv(block)

df2=pd.read\_csv('household\_shortfile.csv', na\_values=['ACORN-U','ACORN-']) # to drop incomplete

df2.dropna(inplace=True)

df\_merged=pd.merge(df1,df2,how='left',on='LCLid')

df\_merged.dropna(inplace=True)

# create plots

# chart1

s1=df\_merged.groupby('Acorn\_grouped',as\_index=True)['KWH/hh (per half hour) '].sum()

chart1=s1.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category")

fig1=chart1.figure

fig1.set\_size\_inches(6,4)

fig1.tight\_layout(pad=2)

chart1.set\_xlabel('Acorn Category')

chart1.set\_ylabel('Electricity Usage KWH')

fig1.savefig(block+' By Acorn Category.png',dpi=125)

# chart2

s2=df\_merged.groupby('Acorn',as\_index=True)['KWH/hh (per half hour) '].sum()

chart2=s2.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Type")

fig2=chart2.figure

fig2.set\_size\_inches(6,4)

fig2.tight\_layout(pad=3)

chart2.set\_xlabel('Acorn Type')

chart2.set\_ylabel('Electricity Usage KWH')

fig2.savefig(block+' By Acorn Type.png',dpi=125)

# char3

s3=df\_merged.groupby(['Acorn\_grouped','Month'])['KWH/hh (per half hour) '].sum()

chart3=s3.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category By Month")

fig3=chart3.figure

fig3.set\_size\_inches(10,6)

fig3.tight\_layout(pad=3)

chart3.set\_xlabel('Acorn Category')

chart3.set\_ylabel('Electricity Usage KWH')

fig3.savefig(block+' By Acorn Category By Month.png',dpi=125)

# chart4

s4=df\_merged.groupby(['Acorn\_grouped','Year'])['KWH/hh (per half hour) '].sum()

chart4=s4.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category By Year")

fig4=chart4.figure

fig4.set\_size\_inches(8,4)

fig4.tight\_layout(pad=3)

chart4.set\_xlabel('Acorn Category')

chart4.set\_ylabel('Electricity Usage KWH')

fig4.savefig(block+' By Acorn Category By Year.png',dpi=125)

# chart5

s5=df\_merged.groupby(['Acorn\_grouped','Hour'])['KWH/hh (per half hour) '].sum()

chart5=s5.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category By Hour")

fig5=chart5.figure

fig5.set\_size\_inches(20,4)

fig5.tight\_layout(pad=3)

chart5.set\_xlabel('Acorn Category')

chart5.set\_ylabel('Electricity Usage KWH')

fig5.savefig(block+' By Acorn Category By Hour.png',dpi=125)

# create database and tables to handle data

def create\_table(): # create table to store summary data for each user LCLid

conn = sqlite3.connect('acorn\_new2.db') # connect (and create it if not exists) to the database

c=conn.cursor()

c.execute('CREATE TABLE IF NOT EXISTS Block\_0(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_1(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_2(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_3(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_4(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_5(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_6(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_7(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_8(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_9(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_10(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_11(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_12(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_13(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Block\_14(LCLid TEXT, Date\_Time DATETIME, KWH\_per\_hh REAL)')

c.execute('CREATE TABLE IF NOT EXISTS Household\_Info(LCLid TEXT, stdorToU TEXT, Acorn TEXT,Acorn\_grouped TEXT, file TEXT)')

c.close() # close cursor

conn.close() # close connection

def Append\_to\_blocktable(block\_filename, tablename): # enter into table to summary data for each user LCLid

conn = sqlite3.connect('acorn\_new2.db') # connect (and create it if not exists) to the database

c=conn.cursor()

blockdata=pd.read\_csv(block\_filename) # figure out a proper format to read and convert the block data file

# write scripts to format blockdata so that each 'data' in 'blockdata' is a string to insert the table

# write scripts to append data to the table

for data in blockdata:

c.execute("INSERT INTO tablename VALUES(data)")

c.close() # close cursor

conn.close() # close connection

**Energy\_Use\_Study\_clean\_format\_data.py**

import pandas as pd

import numpy as np

import matplotlib

import datetime

from matplotlib import pyplot as plt

import project3\_modules as my\_modules

# this code is only to plot inline in Jupyter Notebook

#%matplotlib inline

# data filenames

fileA='informations\_household\_update.csv'

fileB='acorn\_details.csv'

# energy usage files cols = 'LCLid', 'DateTime', 'KWH/hh (per half hour) '

datafile\_list=['block\_0.csv', 'block\_1.csv', 'block\_2.csv', 'block\_3.csv', 'block\_4.csv', 'block\_5.csv', 'block\_6.csv', 'block\_7.csv', 'block\_8.csv', 'block\_9.csv', 'block\_10.csv', 'block\_11.csv', 'block\_12.csv', 'block\_13.csv', 'block\_14.csv']

# convert csv data into DataFrame, convert datetime col from str to DateTime type, add 'Year', 'Month', 'Day\_of\_Week' and 'Hour' col,

# and save each DataFrame as csv file, replacing original csv file.

for item in datafile\_list:

my\_modules.clean\_format\_csv(item)

for filename in datafile\_list:

df=dp.read\_csv(filename)

df.DateTime=pd.to\_datetime(df.DateTime)

df1=df.groupby('LCLid').[DateTime].min()

df1['EndTime']=df.groupby('LCLid').[DateTime].max()

#df['file', 'BeginDateTime', 'EndDateTime']=file,

print filename +": " +str(df.DateTime.min()) + ' through ' + str(df.DateTime.max())

# block\_0.csv file is too big and causes memory problem during data analysis and visualization

# split block\_0.csv file into two block\_0-a.csv and block\_0-b.csv

# final\_file\_list = ['block\_0-a.csv', 'block\_0-b.csv', 'block\_1.csv', 'block\_2.csv', 'block\_3.csv', 'block\_4.csv', 'block\_5.csv', 'block\_6.csv', 'block\_7.csv','block\_8.csv','block\_9.csv','block\_10.csv','block\_11.csv','block\_12.csv','block\_13.csv','block\_14.csv']

df=pd.read\_csv('block\_0.csv')

df1=df[df['LCLid']<'MAC000235'] # splitting by this LCLid gives appro. half the file size

df2=df[df['LCLid']>'MAC000234'] # splitting by this LCLid gives appro. half the file size

df1.to\_csv('block\_0-a', sep=',') # save data to csv format

df2.to\_csv('block\_0-b', sep=',') # save data to csv format

# clean informations\_household\_update.csv

df=pd.read\_csv(fileA, na\_values=[' ','Null','null','N/A','na','',float('inf'), float('-inf'), np.inf, -np.inf])

df.drop\_duplicates(inplace=True)

df.dropna(inplace=True)

df\_household\_info=df[['LCLid','Acorn','Acorn\_grouped']]

df\_household\_info.to\_csv('household\_shortfile.csv', sep=',') #save short household file with cleaned up version

**Energy\_Use\_Study\_Summary.py**

import pandas as pd

import numpy as np

import matplotlib

import datetime

# data filenames

# energy usage files cols = 'LCLid', 'DateTime', 'KWH/hh (per half hour) '

datafile\_list=['block\_0.csv', 'block\_1.csv', 'block\_2.csv', 'block\_3.csv', 'block\_4.csv', 'block\_5.csv', 'block\_6.csv', 'block\_7.csv', 'block\_8.csv', 'block\_9.csv', 'block\_10.csv', 'block\_11.csv', 'block\_12.csv', 'block\_13.csv', 'block\_14.csv']

# create csv lising each smart meter user and mean half hour electricity usage

for filename in datafile\_list:

df=pd.read\_csv(filename)

s1=df.groupby('LCLid')['KWH/hh (per half hour) '].mean()

s1.to\_csv('Mean\_KWH\_perhh\_'+filename)

s2=df.groupby('LCLid')['KWH/hh (per half hour) '].sum()

s2.to\_csv('Total\_KWH\_perhh\_'+filename)

df.DateTime=pd.to\_datetime(df.DateTime)

s3=df.groupby('LCLid')['DateTime'].min()

s3.to\_csv('Start\_'+filename)

s4=df.groupby('LCLid')['DateTime'].max()

s4.to\_csv('End\_'+filename)

for filename in datafile\_list:

file1='Start\_'+filename

file2='End\_'+filename

file3='Total\_KWH\_perhh\_'+filename

file4='Mean\_KWH\_perhh\_'+filename

df1=pd.read\_csv(file1, names=['LCLid','Start'])

df2=pd.read\_csv(file2, names=['LCLid','End'])

df3=pd.read\_csv(file3, names=['LCLid','Total\_Usage\_KWH'])

df4=pd.read\_csv(file4, names=['LCLid','Mean\_Usage\_KWH/hh'])

df5=pd.merge(left=df1, right=df2, how='left', left\_on='LCLid',right\_on='LCLid')

df5=pd.merge(left=df5, right=df3, how='left', left\_on='LCLid',right\_on='LCLid')

df5=pd.merge(left=df5, right=df4, how='left', left\_on='LCLid',right\_on='LCLid')

df5.to\_csv('summary\_'+filename)

print filename

datafile\_list=['block\_0.csv', 'block\_1.csv', 'block\_2.csv', 'block\_3.csv', 'block\_4.csv', 'block\_5.csv', 'block\_6.csv', 'block\_7.csv', 'block\_8.csv', 'block\_9.csv', 'block\_10.csv', 'block\_11.csv', 'block\_12.csv', 'block\_13.csv', 'block\_14.csv']

df6=pd.read\_csv('household\_shortfile.csv') # to read and merge acorn type info

for filename in datafile\_list:

df7=pd.read\_csv('summary\_'+filename) #names=['rows','LCLid','Start','End','Total\_Usage\_KWH','Mean\_Usage\_KWH/hh'])

df7=pd.merge(left=df7, right=df6, how='left', left\_on='LCLid', right\_on='LCLid')

df7.to\_csv('summary\_with\_household\_info\_'+filename)

# compile final summary file

new\_datafile\_list=['block\_1.csv', 'block\_2.csv', 'block\_3.csv', 'block\_4.csv', 'block\_5.csv', 'block\_6.csv', 'block\_7.csv', 'block\_8.csv', 'block\_9.csv', 'block\_10.csv', 'block\_11.csv', 'block\_12.csv', 'block\_13.csv', 'block\_14.csv']

df8=pd.read\_csv('summary\_with\_household\_info\_block\_0.csv')

for filename in new\_datafile\_list:

df9=pd.read\_csv('summary\_with\_household\_info\_'+filename)

df8=pd.concat([df8,df9],axis=0)

df8.dropna(inplace=True)

df8.to\_csv('summary\_with\_household\_info\_all\_blocks.csv')

# Compile summary data file

df10=pd.read\_csv('final\_summary\_all\_blocks.csv')

Total\_Acorn\_Category=df10.groupby('Acorn\_grouped')['Total\_Usage\_KWH'].sum()

Total\_Acorn\_Category.to\_csv('Overall\_Usage\_By\_Category.csv') # save total usage KWH data by Acorn Category to csv

Total\_Acorn\_Type=df10.groupby('Acorn')['Total\_Usage\_KWH'].sum()

Total\_Acorn\_Type.to\_csv('Overall\_Usage\_By\_Type.csv') # save total usage KWH data by Acorn type to csv

**Energy\_Use\_Study\_Analyze\_Visualize.py**

import pandas as pd

import numpy as np

import matplotlib

from matplotlib import pyplot as plt

# this code is only to plot inline in Jupyter Notebook

#%matplotlib inline

# energy usage files cols = LCLid, DateTime, KWH/hh (per half hour)

datafile\_list=['block\_0.csv','block\_1.csv','block\_2.csv','block\_3.csv','block\_4.csv','block\_5.csv','block\_6.csv',

'block\_7.csv','block\_8.csv','block\_9.csv','block\_10.csv','block\_11.csv','block\_12.csv','block\_13.csv','block\_14.csv']

# block\_0.csv file is too big and causes memory problem during data analysis and visualization

# split block\_0.csv file into two block\_0-a.csv and block\_0-b.csv

# final\_file\_list = ['block\_0-a.csv', 'block\_0-b.csv', 'block\_1.csv', 'block\_2.csv', 'block\_3.csv', 'block\_4.csv', 'block\_5.csv', 'block\_6.csv', 'block\_7.csv','block\_8.csv','block\_9.csv','block\_10.csv','block\_11.csv','block\_12.csv','block\_13.csv','block\_14.csv']

# analyze and visualize data one block at a time and merge dfs

# enter one filename at a time and run this scripts

block='block\_0-b.csv'

df1=pd.read\_csv(block)

df2=pd.read\_csv('household\_shortfile.csv', na\_values=['ACORN-U','ACORN-']) # to drop incomplete entries

df2.dropna(inplace=True)

df\_merged=pd.merge(left=df1,right=df2,how='left',on='LCLid')

df\_merged.dropna(inplace=True)

# create plots

s1=df\_merged.groupby('Acorn\_grouped',as\_index=True)['KWH/hh (per half hour) '].sum()

chart1=s1.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category")

fig1=chart1.figure

fig1.set\_size\_inches(6,4)

fig1.tight\_layout(pad=2)

chart1.set\_xlabel('Acorn Category')

chart1.set\_ylabel('Electricity Usage KWH')

fig1.savefig(block+' By Acorn Category.png',dpi=125)

s2=df\_merged.groupby('Acorn',as\_index=True)['KWH/hh (per half hour) '].sum()/1000000

chart2=s2.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Type")

fig2=chart2.figure

fig2.set\_size\_inches(6,4)

fig2.tight\_layout(pad=3)

chart2.set\_xlabel('Acorn Type')

chart2.set\_ylabel('Electricity Usage, x 1,000,000 KWH')

fig2.savefig(block+' By Acorn Type.png',dpi=125)

s3=df\_merged.groupby(['Acorn\_grouped','Month'])['KWH/hh (per half hour) '].sum()/1000000

chart3=s3.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category By Month")

fig3=chart3.figure

fig3.set\_size\_inches(20,8)

fig3.tight\_layout(pad=2)

chart3.set\_xlabel('Acorn Category by Month')

chart3.set\_ylabel('Electricity Usage, x 1,000,000 KWH')

fig3.savefig('NEW\_'+block+' By Acorn Category By Month.png',dpi=125)

s4=df\_merged.groupby(['Acorn\_grouped','Year'])['KWH/hh (per half hour) '].sum()

chart4=s4.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category By Year")

fig4=chart4.figure

fig4.set\_size\_inches(8,4)

fig4.tight\_layout(pad=3)

chart4.set\_xlabel('Acorn Category')

chart4.set\_ylabel('Electricity Usage KWH')

fig4.savefig(block+' By Acorn Category By Year.png',dpi=125)

'''

s5=df\_merged.groupby(['Acorn\_grouped','Hour'])['KWH/hh (per half hour) '].sum()/1000000

chart5=s5.plot(kind='bar',title=block+" Total Electricity Usage By Acorn Category By Hour")

fig5=chart5.figure

fig5.set\_size\_inches(20,6)

fig5.tight\_layout(pad=2)

chart5.set\_xlabel('Acorn Category by Hour')

chart5.set\_ylabel('Electricity Usage, x 1,000,000 KWH')

fig5.savefig('NEW\_'+block+' By Acorn Category By Hour.png',dpi=125)

'''

# create chart for final overall usage by Acorn Type

df6=pd.read\_csv('Overall\_Usage\_By\_Type.csv', names=['Type', 'Total\_Usage'])

chart6=df6.plot(legend=False,x='Type', y='Total\_Usage',kind='bar',title="Total Electricity Usage By Acorn Type")

fig6=chart6.figure

fig6.set\_size\_inches(8,4)

fig6.tight\_layout(pad=3)

chart6.set\_xlabel('Acorn Type')

chart6.set\_ylabel('Electricity Usage KWH')

fig6.savefig('Total Usage By Acorn Type.png',dpi=125)

# create chart for final overall usage by Acorn Category

df7=pd.read\_csv('Overall\_Usage\_By\_Category.csv', names=['Acorn Category', 'Total\_Usage'])

df7.Total\_Usage=df7.Total\_Usage/1000000

chart7=df7.plot(legend=False,x='Acorn Category', y='Total\_Usage',kind='bar',title="Total Electricity Usage By Acorn Category")

fig7=chart7.figure

fig7.set\_size\_inches(6,5)

fig7.tight\_layout(pad=4)

chart7.set\_xlabel('Acorn Category')

chart7.set\_ylabel('Electricity Usage, x 1,000,000 KWH')

fig7.savefig('Total Usage By Acorn Category.png',dpi=125)

# API weather data from darksky.net for UK locations

# APPENDIX B: Charts

# Figures 1 - 16: Total Electricity Usage by Acorn Category Per Data Block

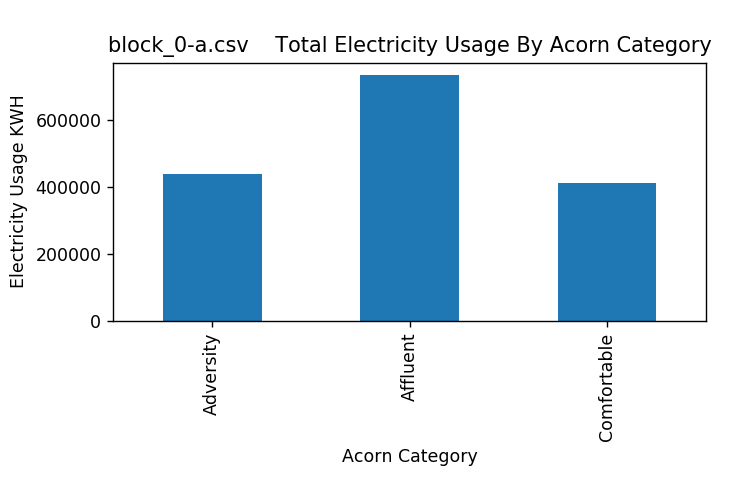


Figure 1: Total Electricity Usage by Acorn Category – Block\_0-a

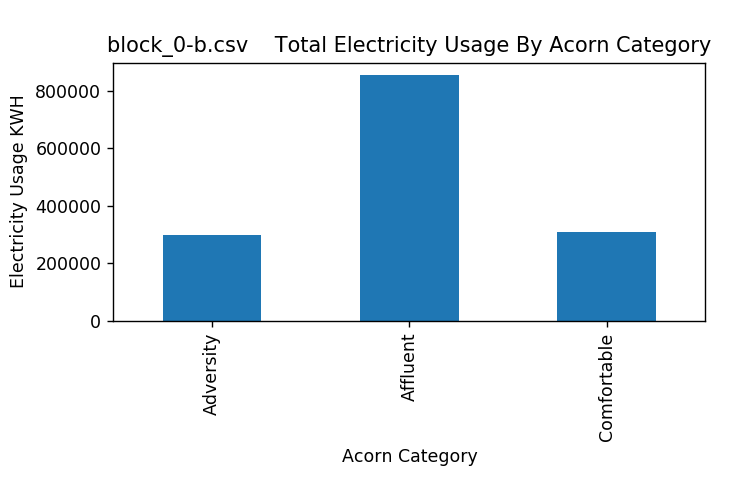


Figure 2: Total Electricity Usage by Acorn Category – Block\_0-b

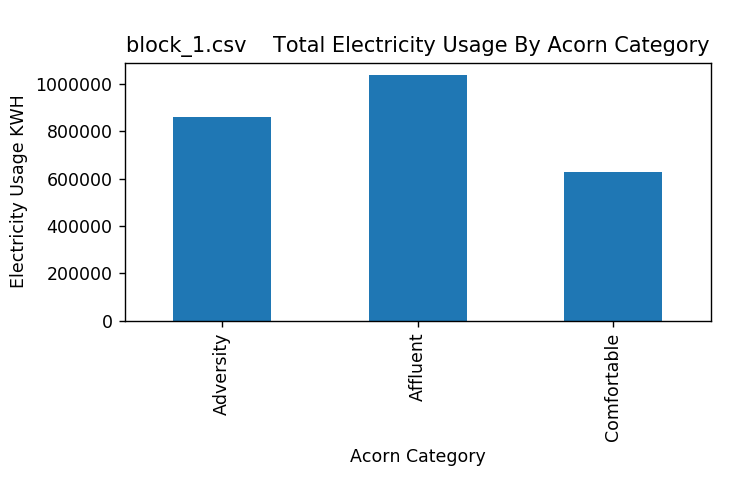


Figure 3: Total Electricity Usage by Acorn Category – Block\_1

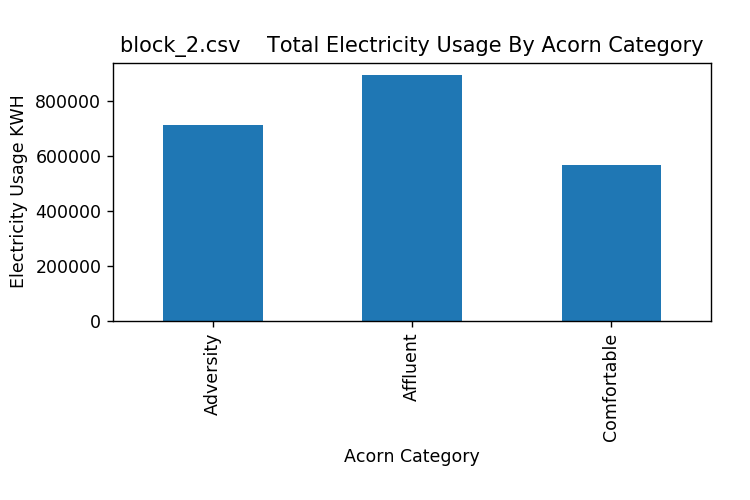


Figure 4: Total Electricity Usage by Acorn Category – Block\_2

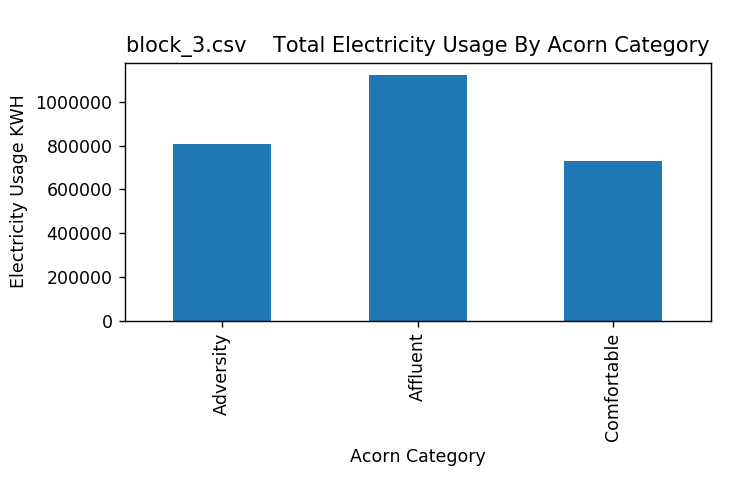


Figure 5: Total Electricity Usage by Acorn Category – Block\_3

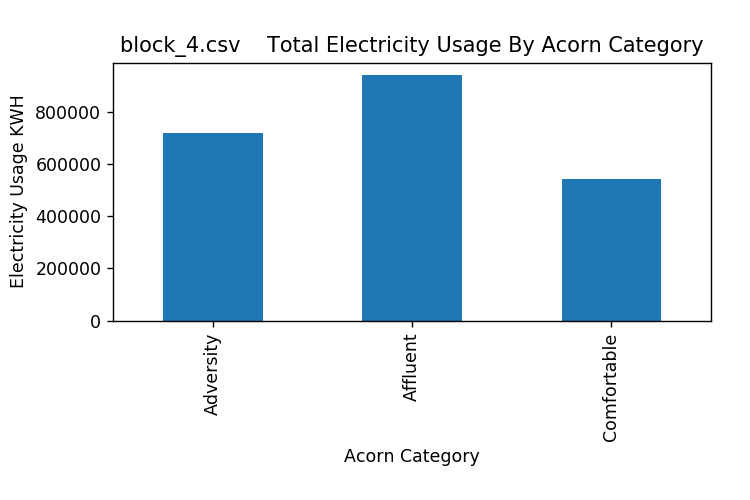


Figure 6: Total Electricity Usage by Acorn Category – Block\_4

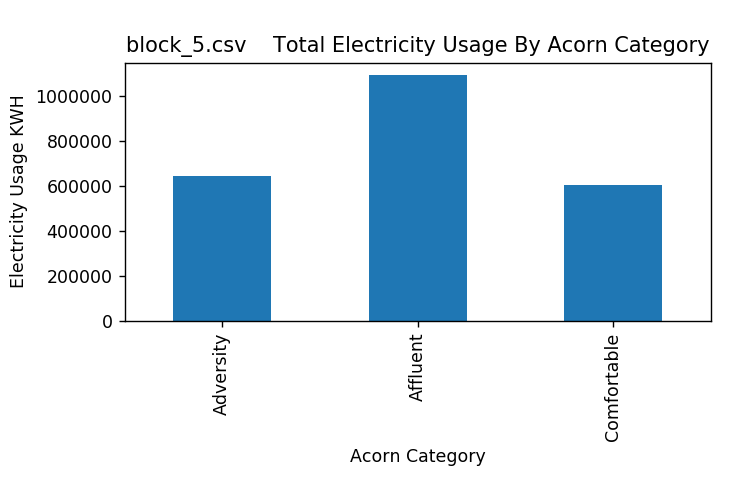


Figure 7: Total Electricity Usage by Acorn Category – Block\_5

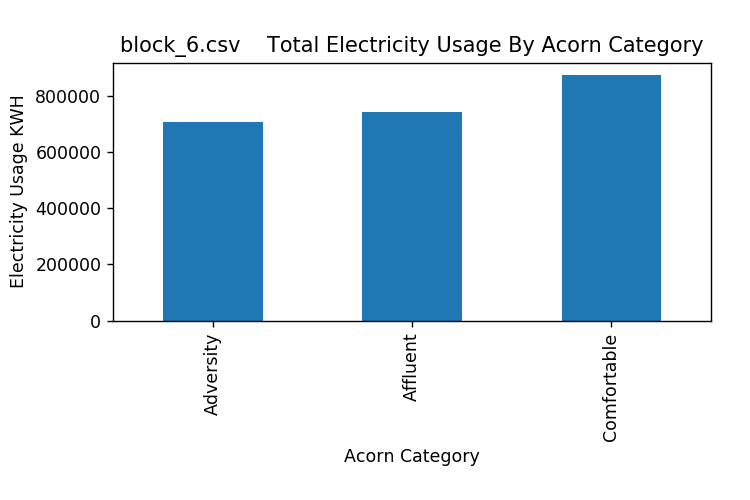


Figure 8: Total Electricity Usage by Acorn Category – Block\_6

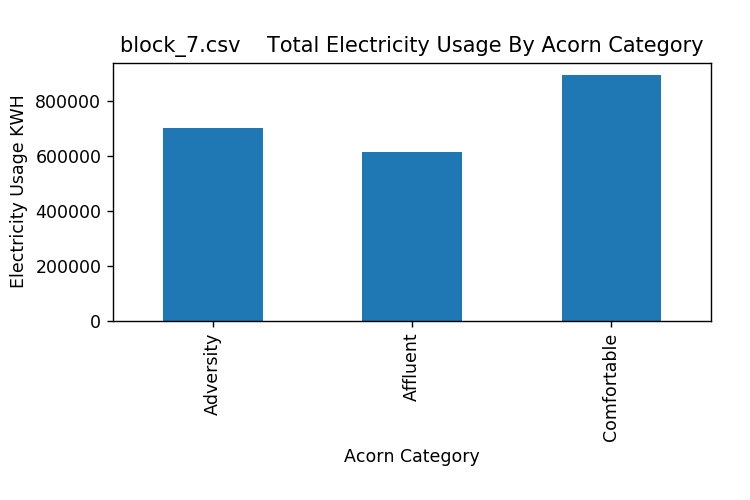


Figure 9: Total Electricity Usage by Acorn Category – Block\_7

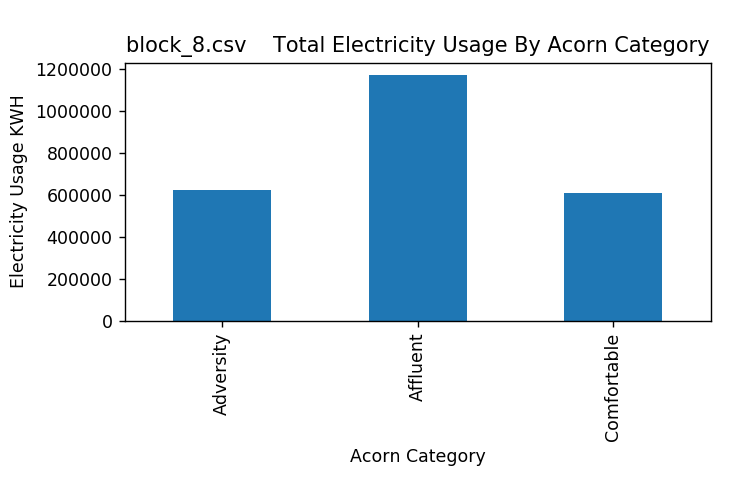


Figure 10: Total Electricity Usage by Acorn Category – Block\_8

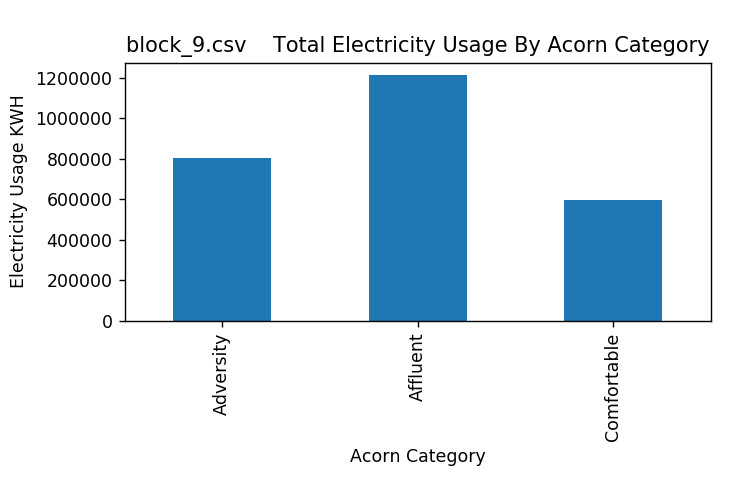


Figure 11: Total Electricity Usage by Acorn Category – Block\_9

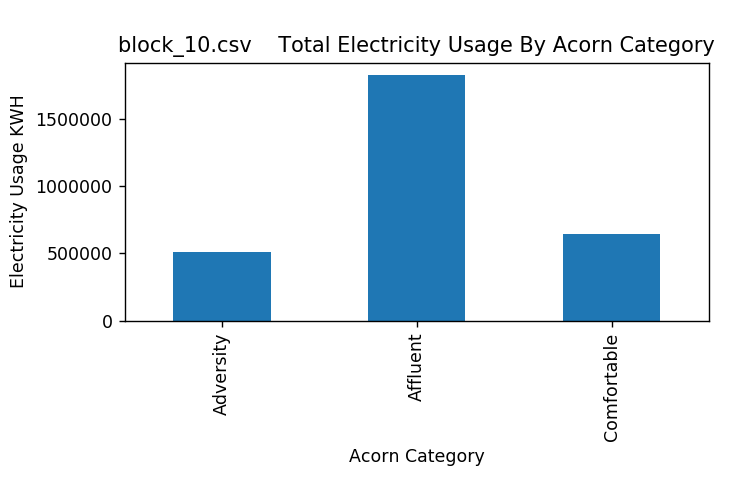


Figure 12: Total Electricity Usage by Acorn Category – Block\_10

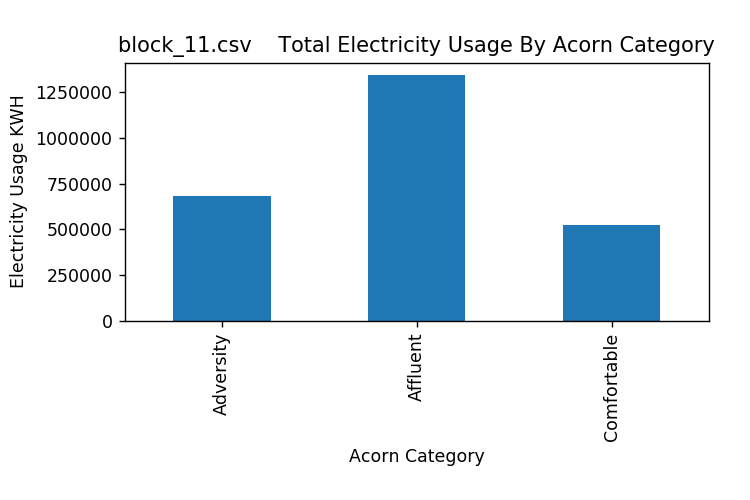


Figure 13: Total Electricity Usage by Acorn Category – Block\_11

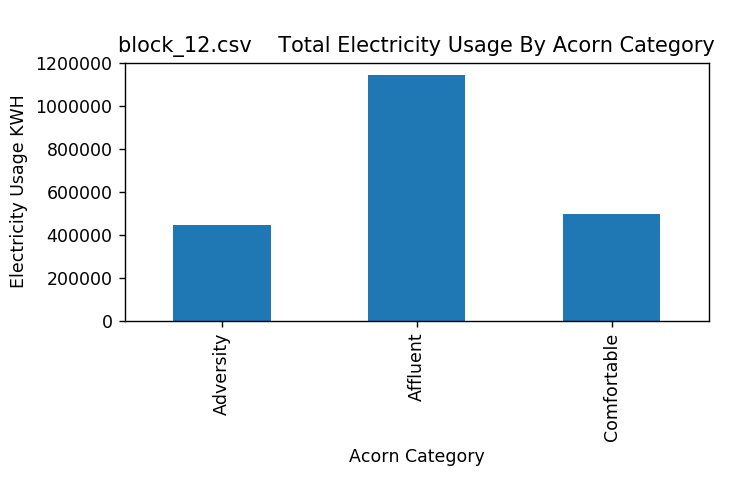


Figure 14: Total Electricity Usage by Acorn Category – Block\_12

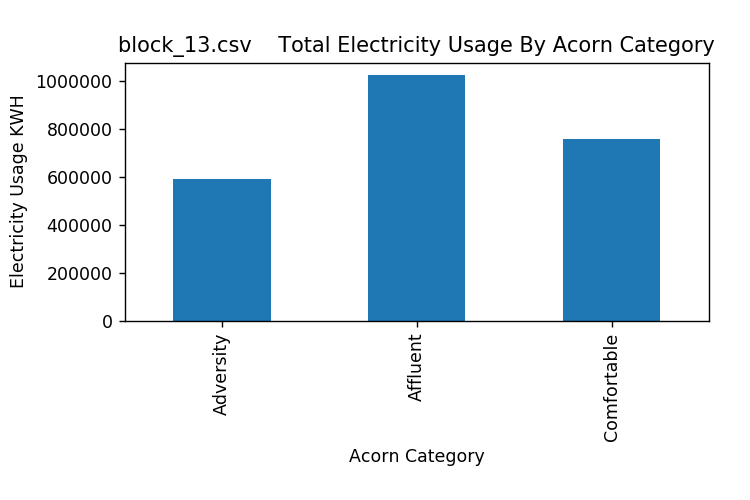


Figure 15: Total Electricity Usage by Acorn Category – Block\_13

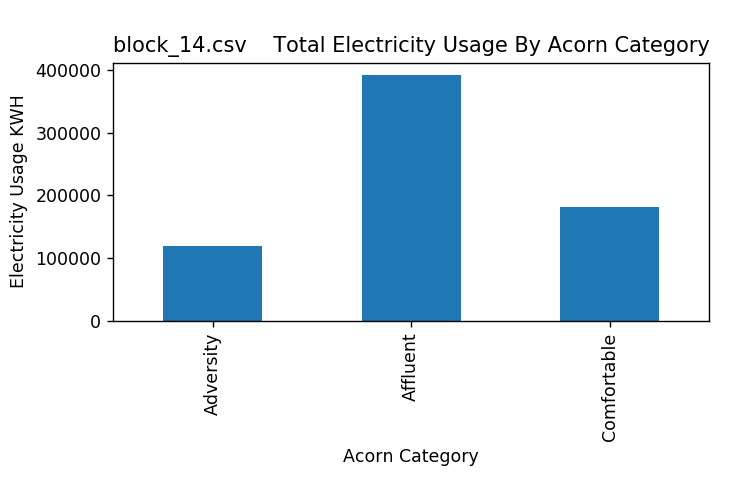


Figure 16: Total Electricity Usage by Acorn Category – Block\_14

# Figures 17 – 32: Total Electricity Usage by Acorn Category by Month Per Data Block

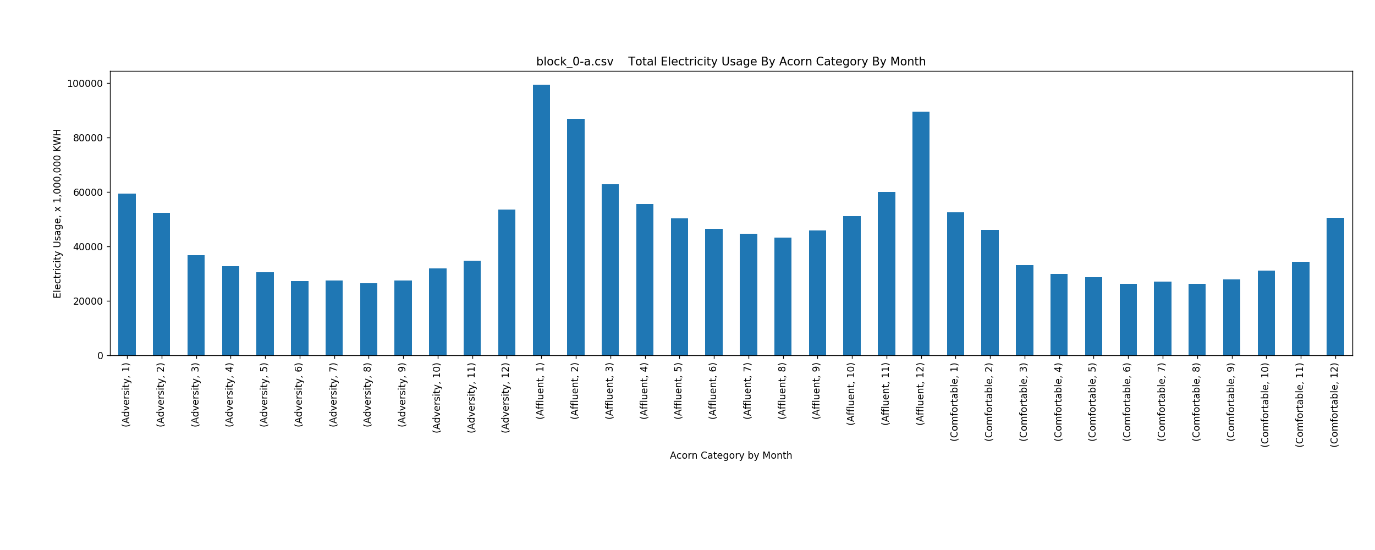


Figure 17: Total Electricity Usage by Acorn Category by Month – Block\_0-a

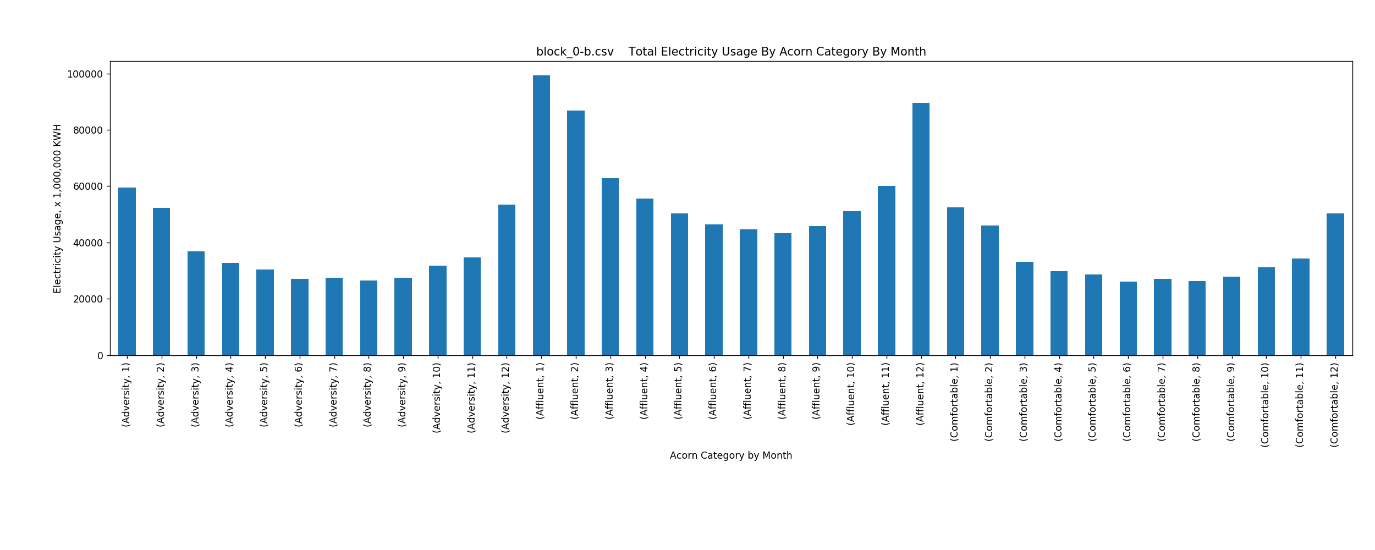


Figure 18: Total Electricity Usage by Acorn Category by Month – Block\_0-b

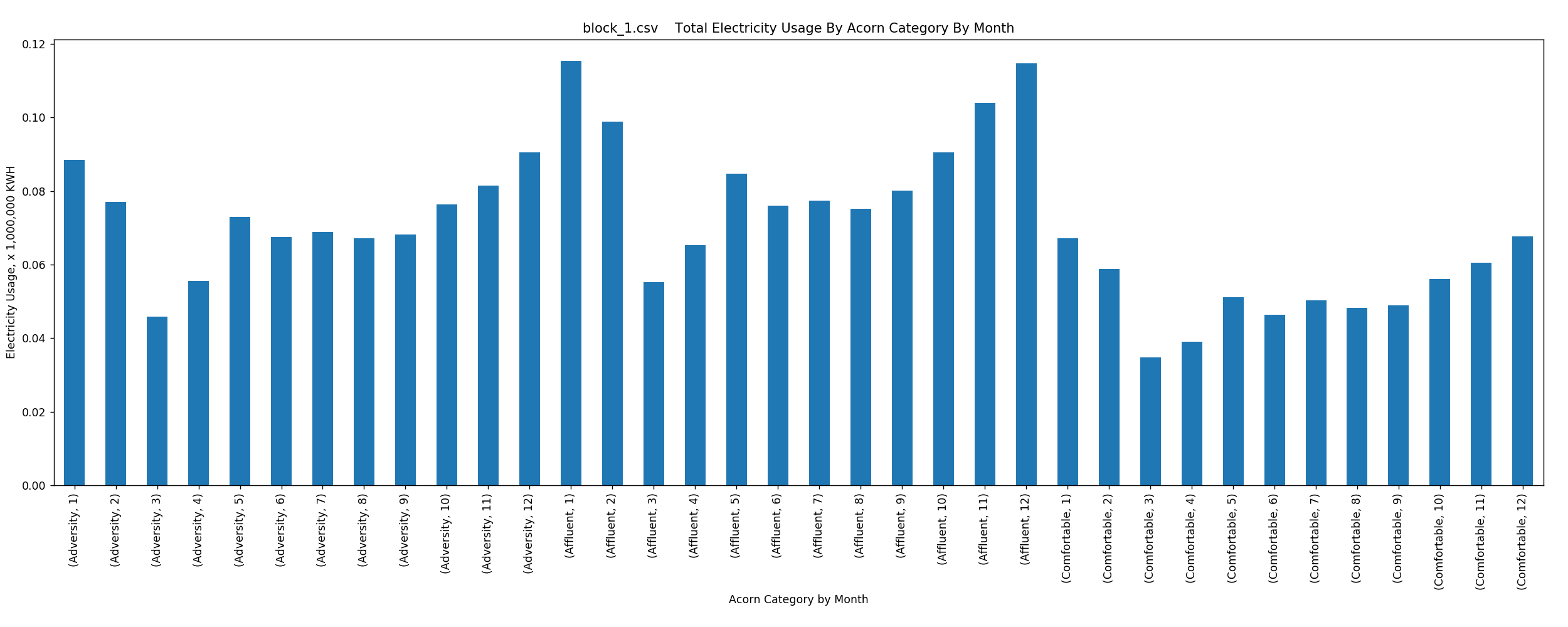


Figure 19: Total Electricity Usage by Acorn Category by Month – Block\_1

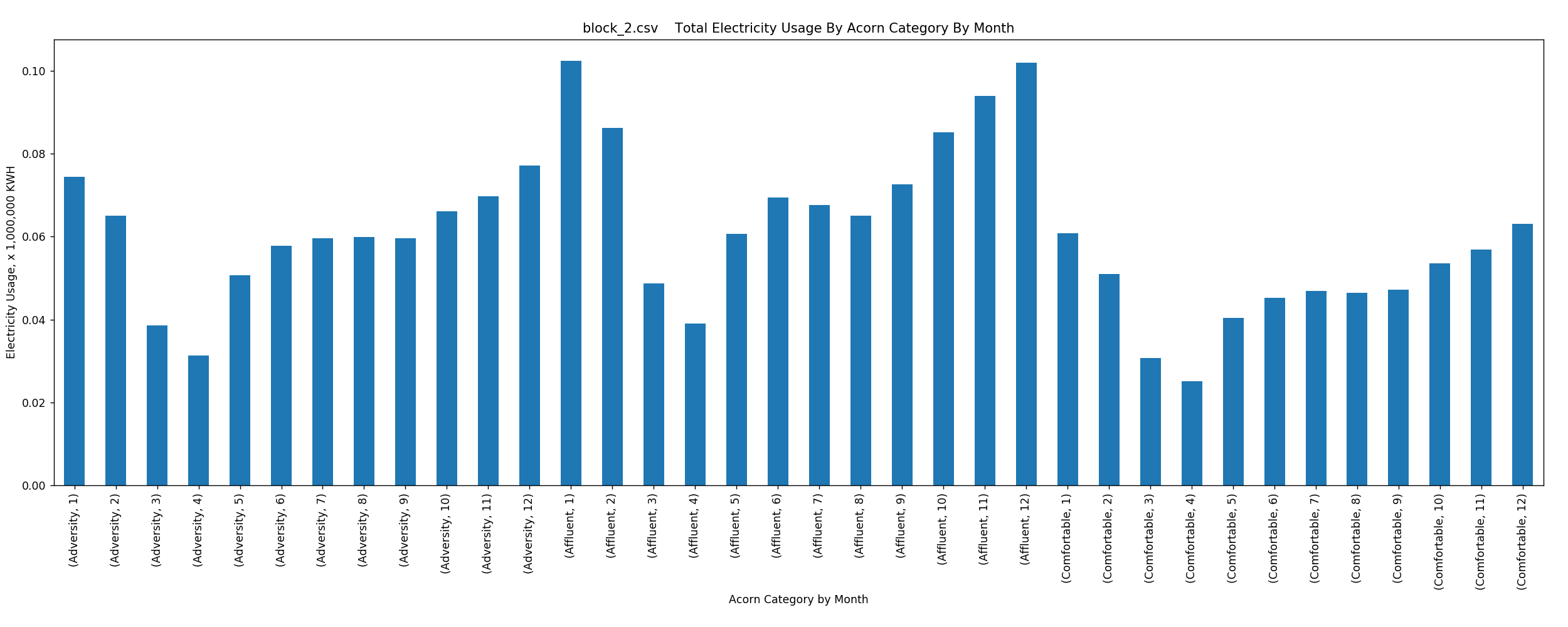


Figure 20: Total Electricity Usage by Acorn Category by Month – Block\_2

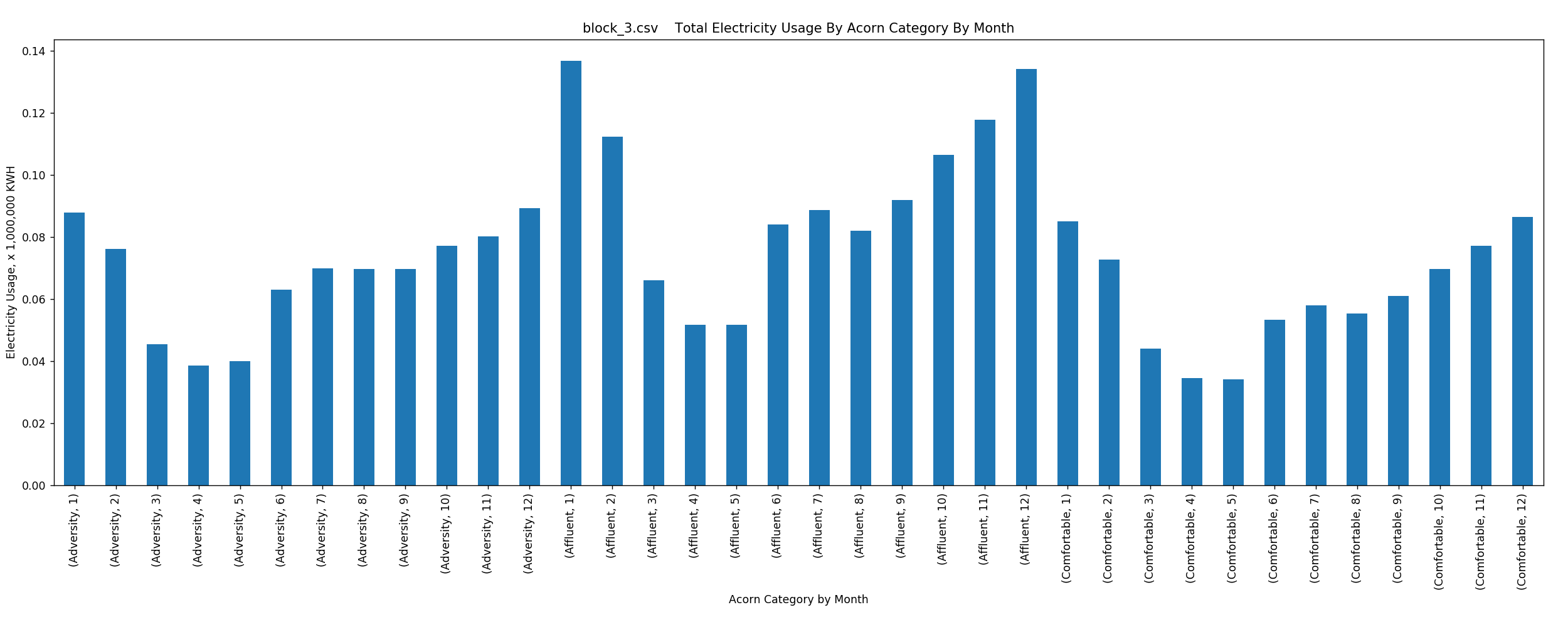


Figure 21: Total Electricity Usage by Acorn Category by Month – Block\_3

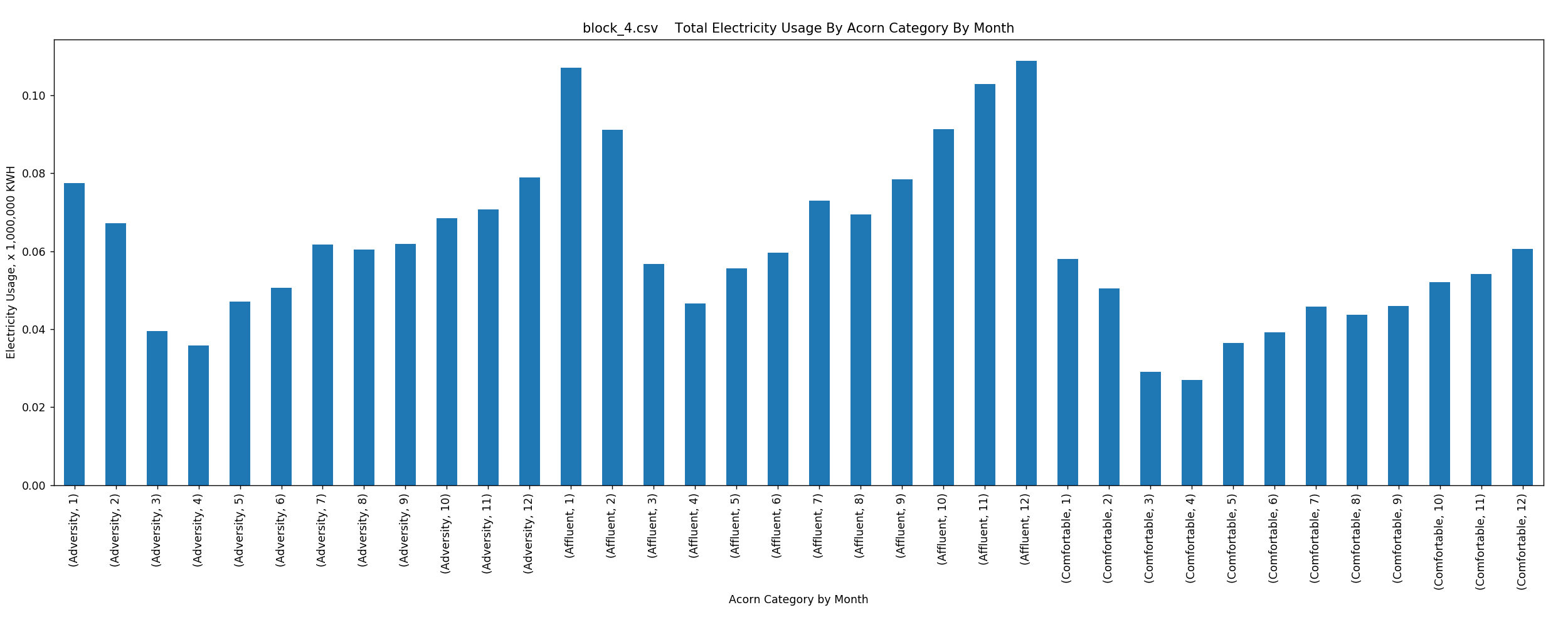


Figure 22: Total Electricity Usage by Acorn Category by Month – Block\_4

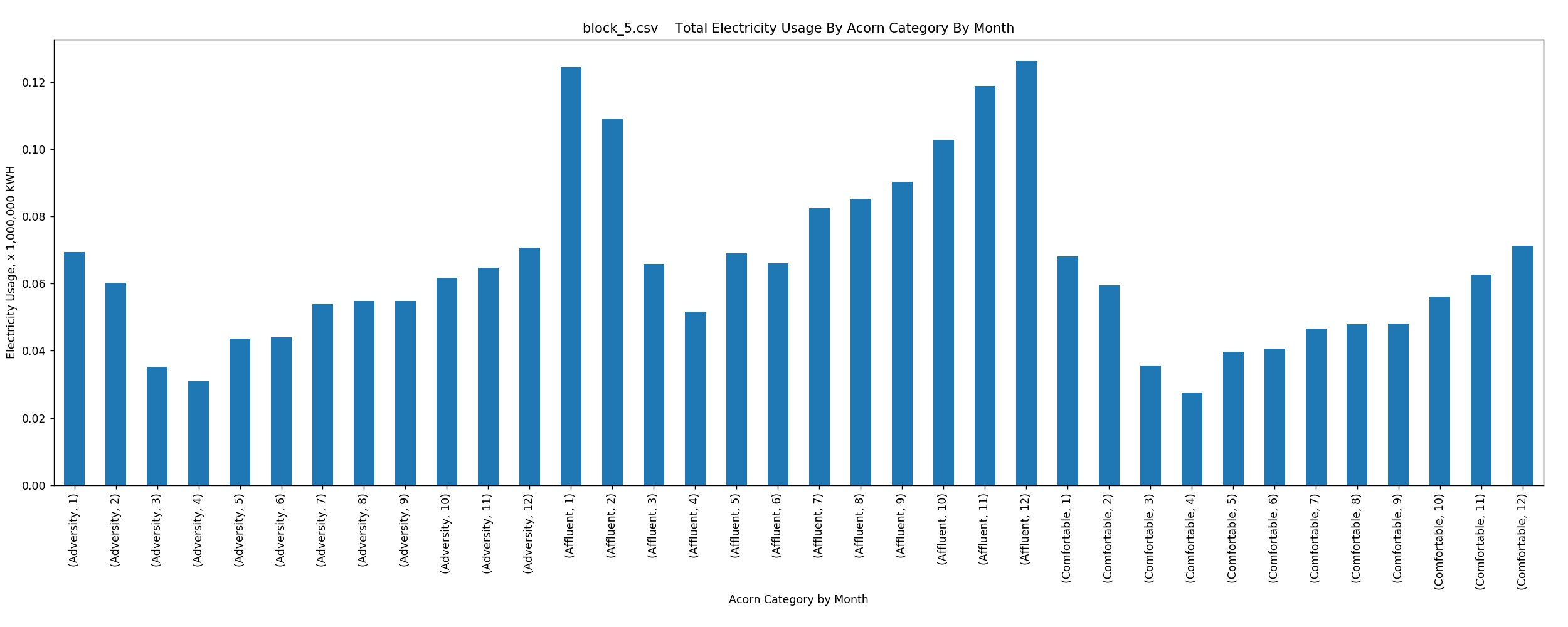


Figure 23: Total Electricity Usage by Acorn Category by Month – Block\_5

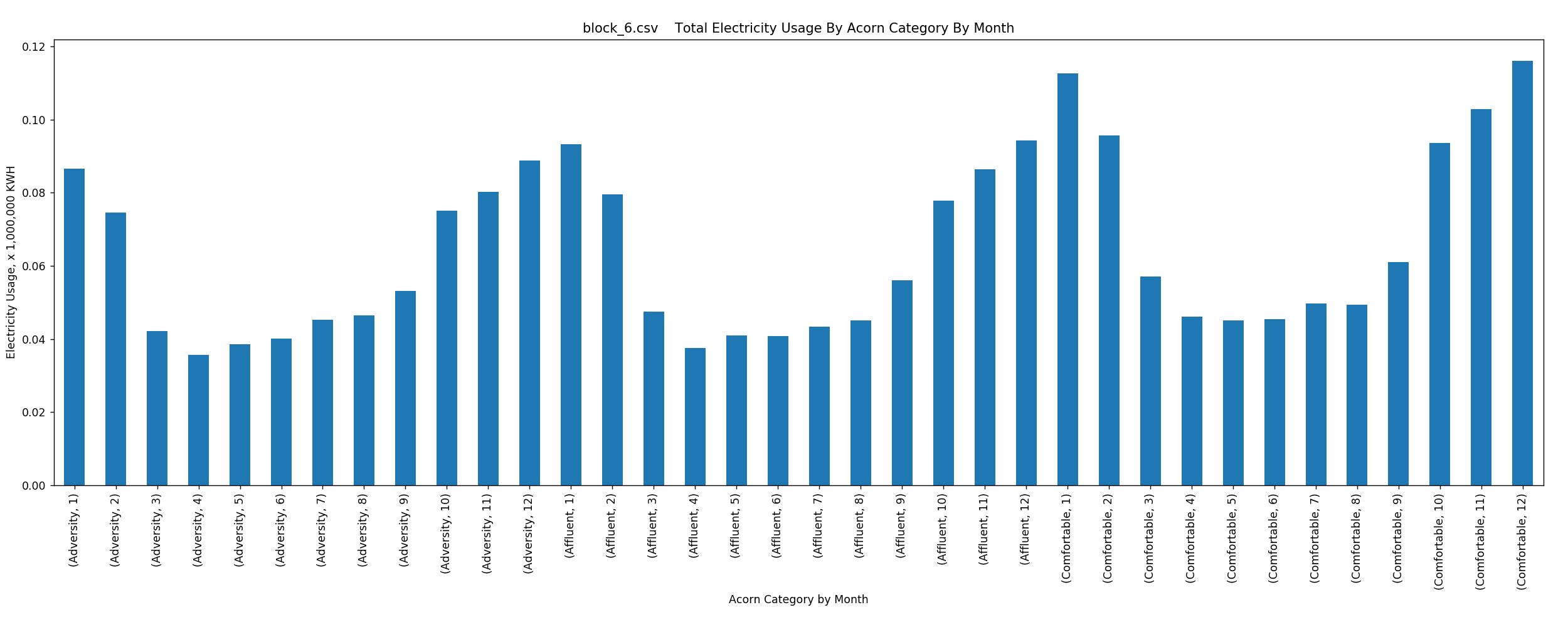


Figure 24: Total Electricity Usage by Acorn Category by Month – Block\_6

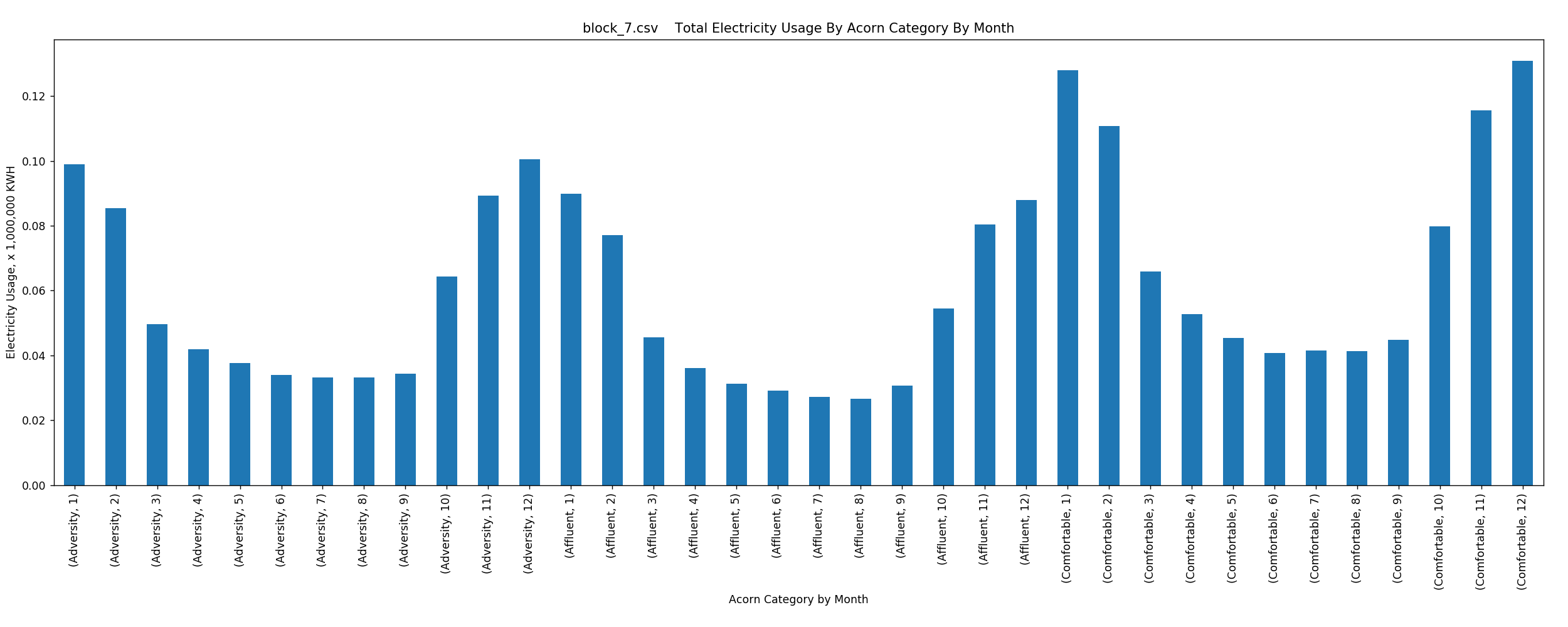


Figure 25: Total Electricity Usage by Acorn Category by Month – Block\_7

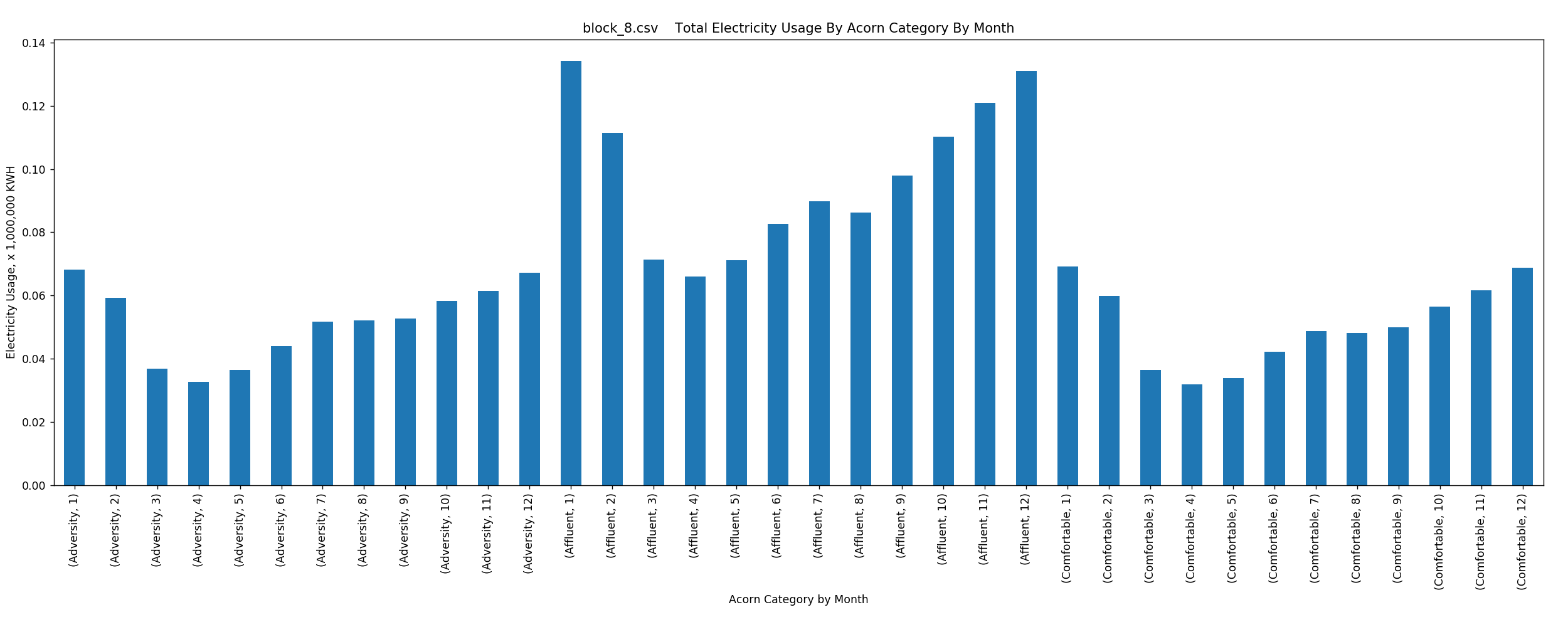


Figure 26: Total Electricity Usage by Acorn Category by Month – Block\_8

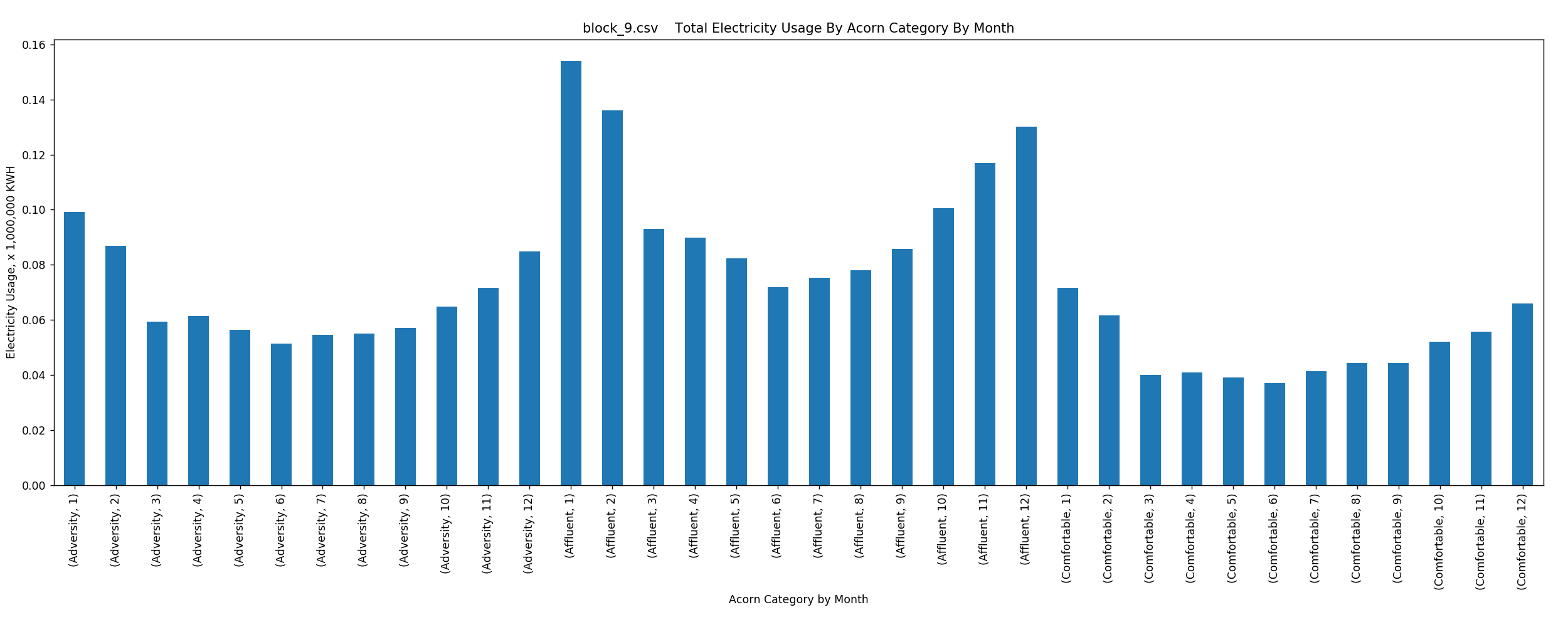


Figure 27: Total Electricity Usage by Acorn Category by Month – Block\_9

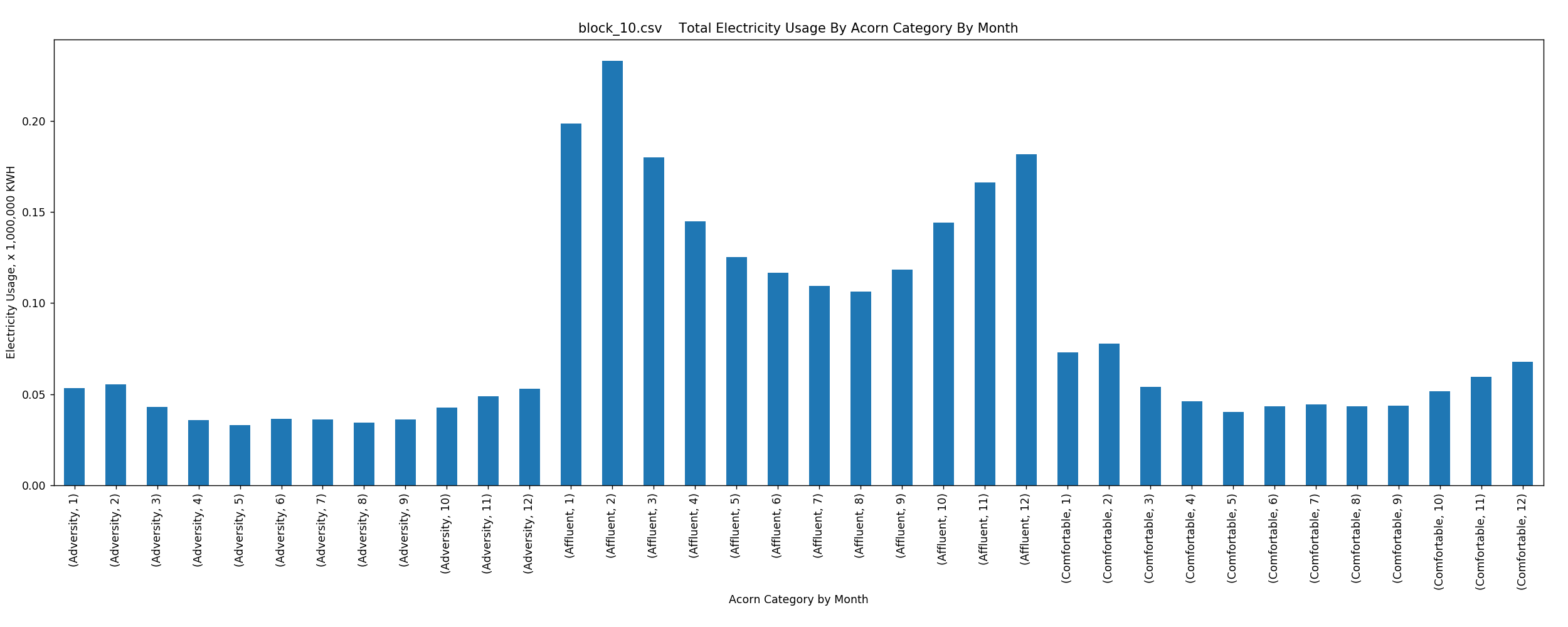


Figure 28: Total Electricity Usage by Acorn Category by Month – Block\_10

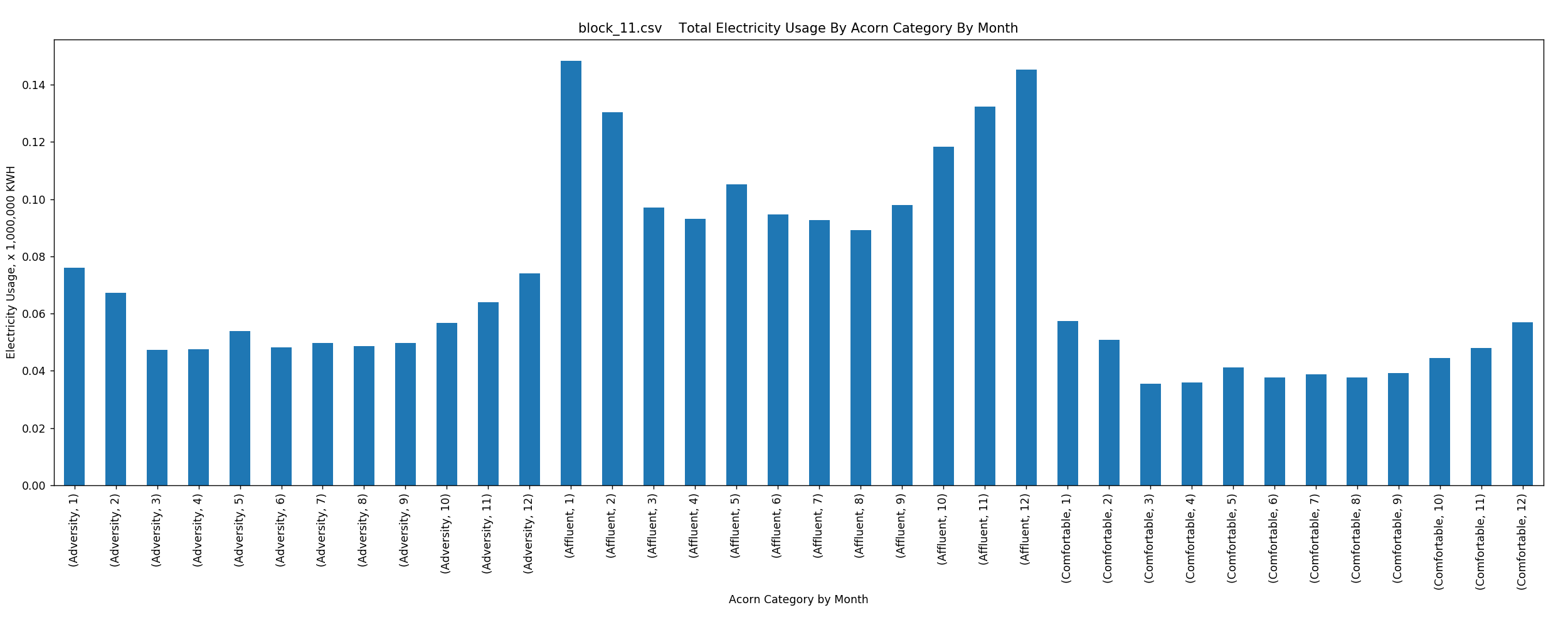


Figure 29: Total Electricity Usage by Acorn Category by Month – Block\_11

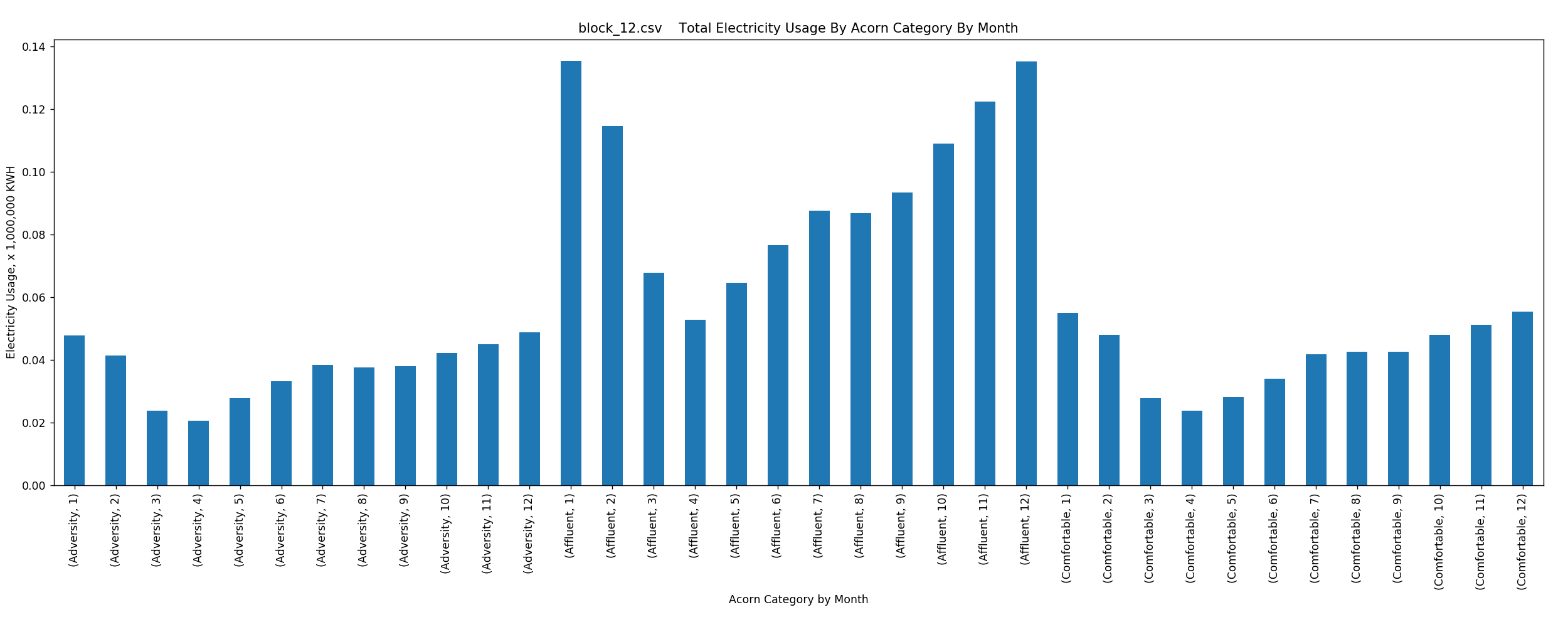


Figure 30: Total Electricity Usage by Acorn Category by Month – Block\_12

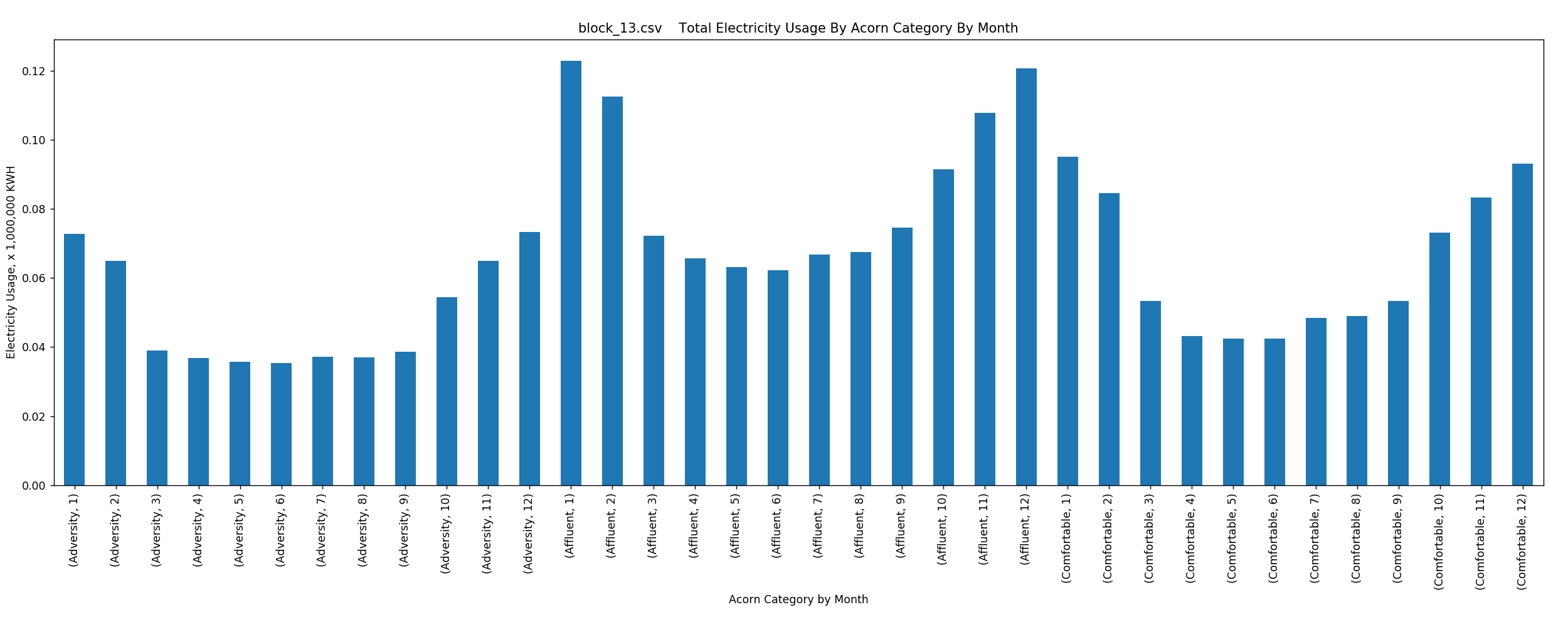


Figure 31: Total Electricity Usage by Acorn Category by Month – Block\_13

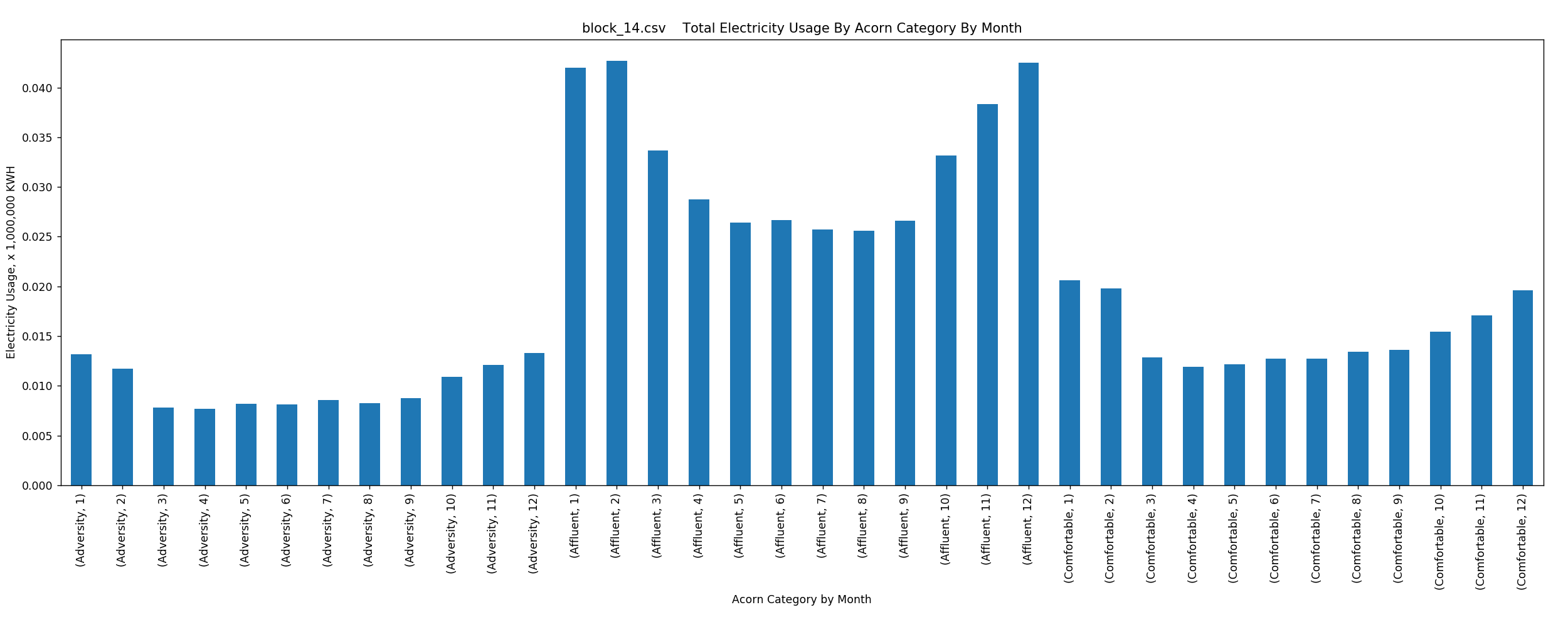


Figure 32: Total Electricity Usage by Acorn Category by Month – Block\_14

# Figures 33 – 48: Total Electricity Usage by Acorn Type Per Data Block

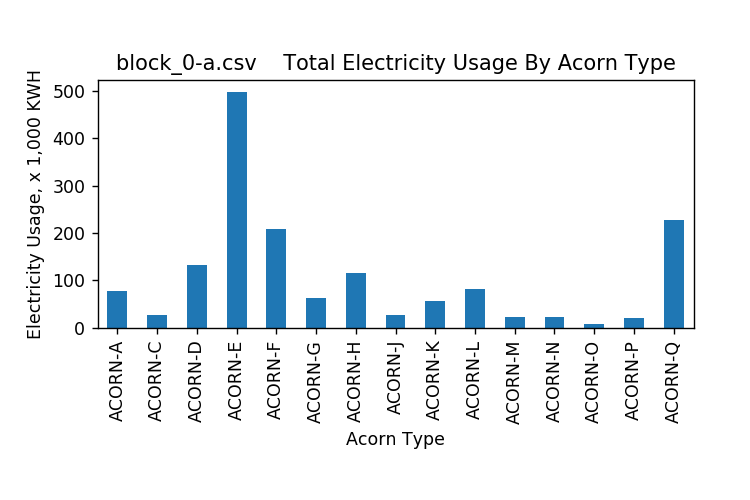


Figure 33: Total Electricity Usage by Acorn Type – Block\_0-a

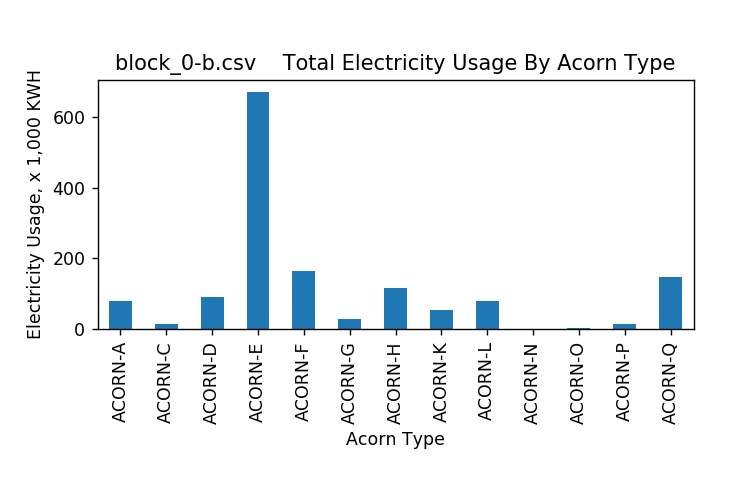


Figure 34: Total Electricity Usage by Acorn Type – Block\_0-b

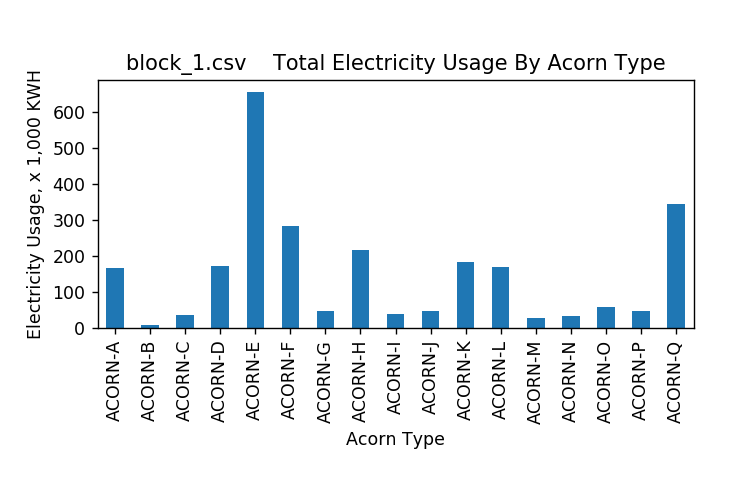


Figure 35: Total Electricity Usage by Acorn Type – Block\_1

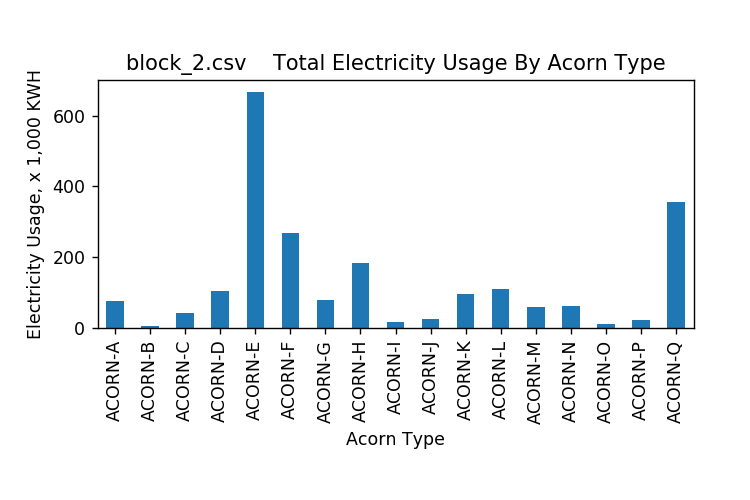


Figure 36: Total Electricity Usage by Acorn Type – Block\_2

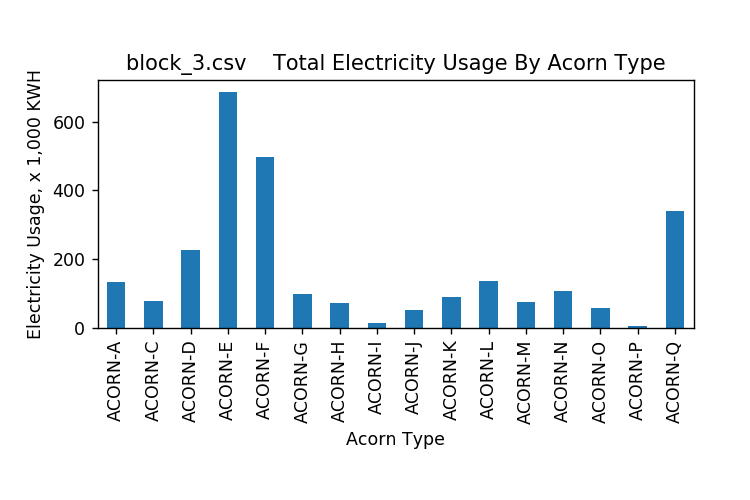


Figure 37: Total Electricity Usage by Acorn Type – Block\_3

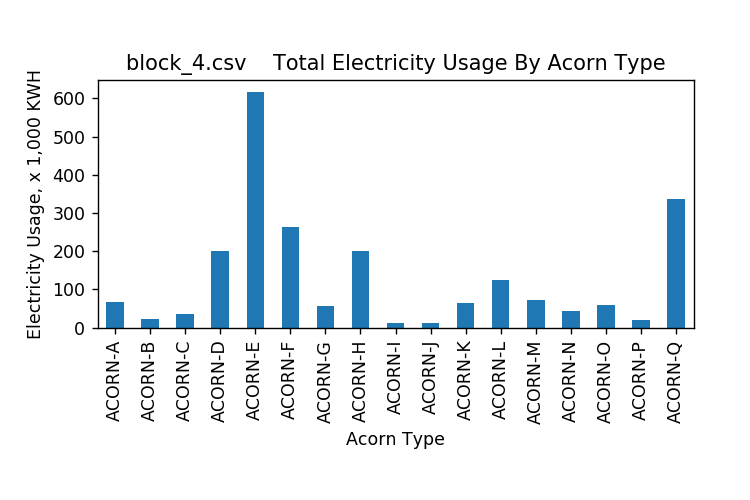


Figure 38: Total Electricity Usage by Acorn Type – Block\_4

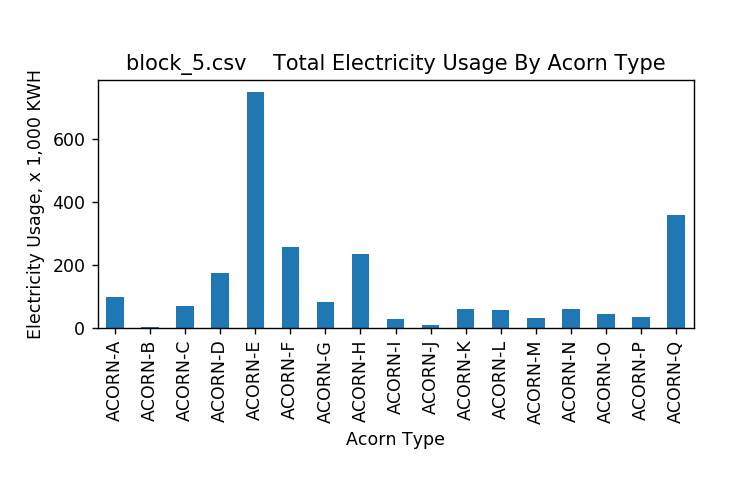


Figure 39: Total Electricity Usage by Acorn Type – Block\_5

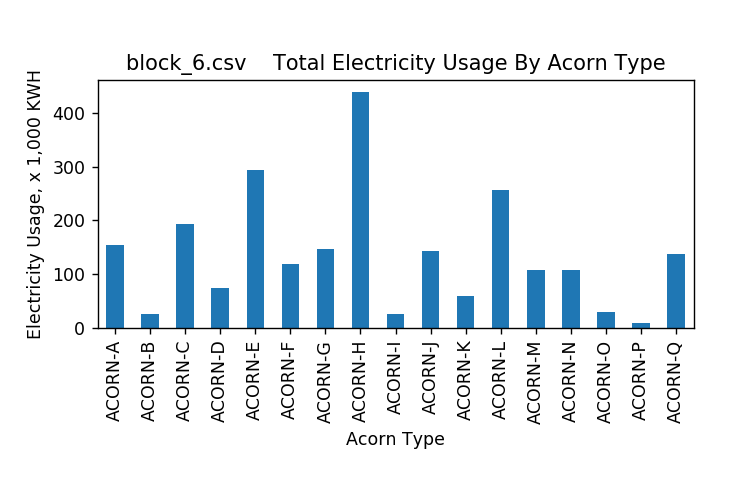


Figure 40: Total Electricity Usage by Acorn Type – Block\_6

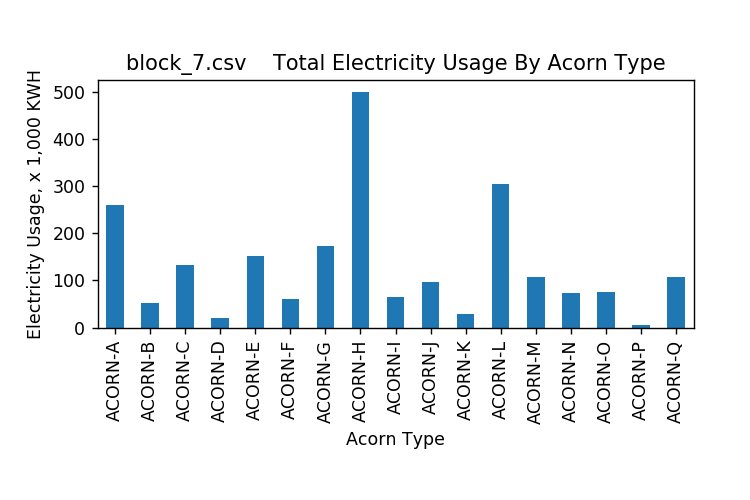


Figure 41: Total Electricity Usage by Acorn Type – Block\_7

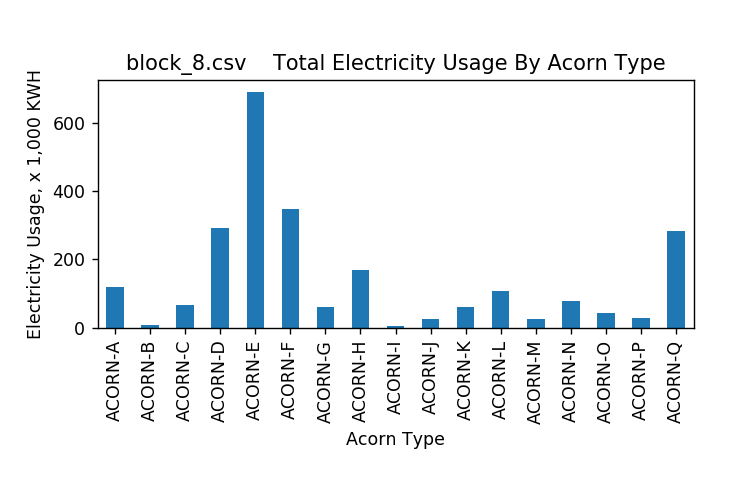


Figure 42: Total Electricity Usage by Acorn Type – Block\_8

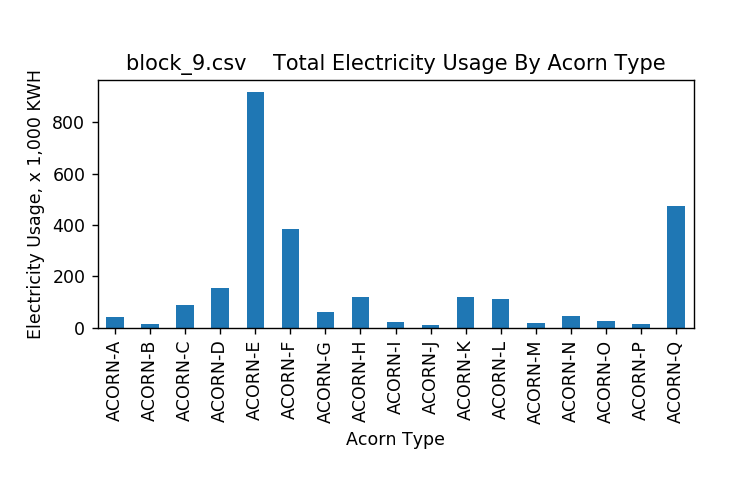


Figure 43: Total Electricity Usage by Acorn Type – Block\_9

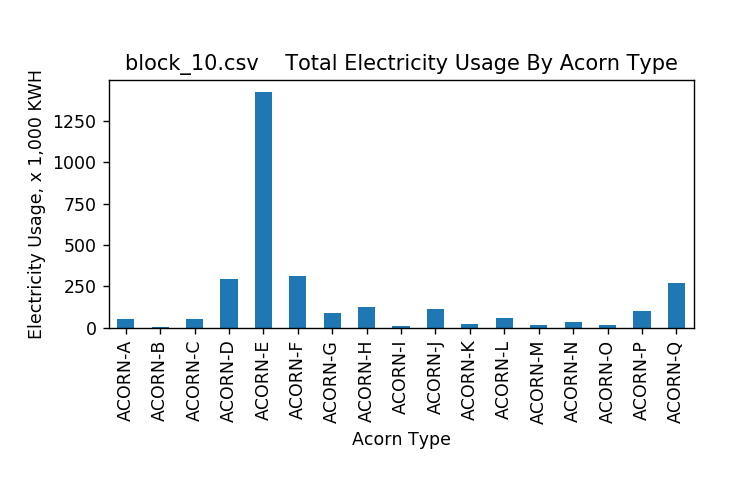


Figure 44: Total Electricity Usage by Acorn Type – Block\_10

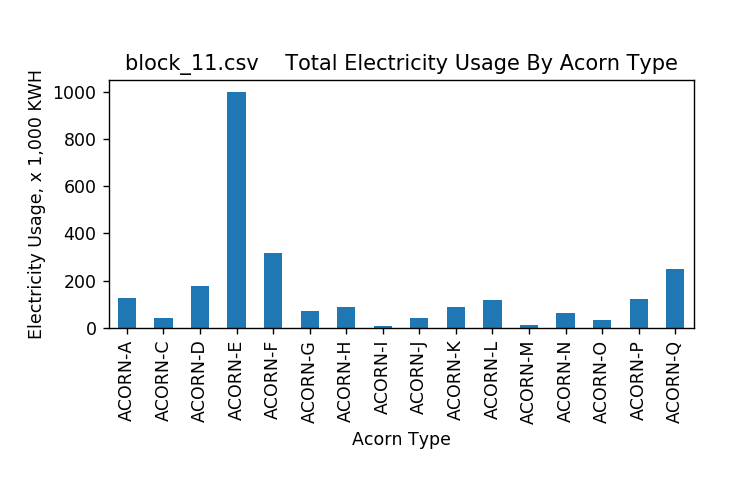


Figure 45: Total Electricity Usage by Acorn Type – Block\_11

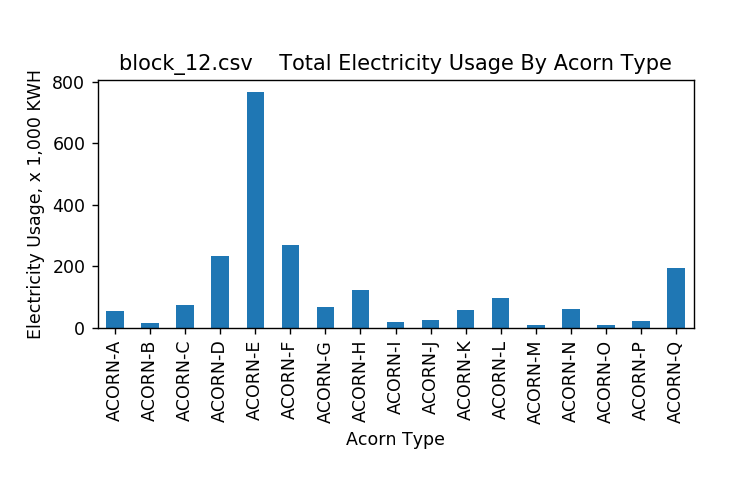


Figure 46: Total Electricity Usage by Acorn Type – Block\_12

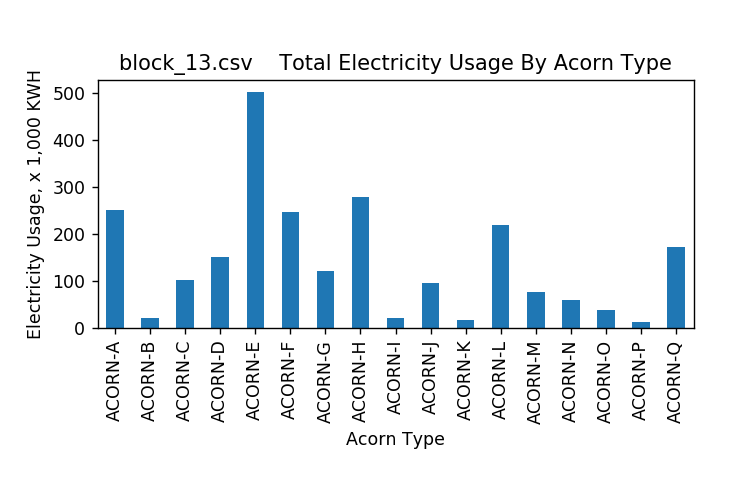


Figure 47: Total Electricity Usage by Acorn Type – Block\_13

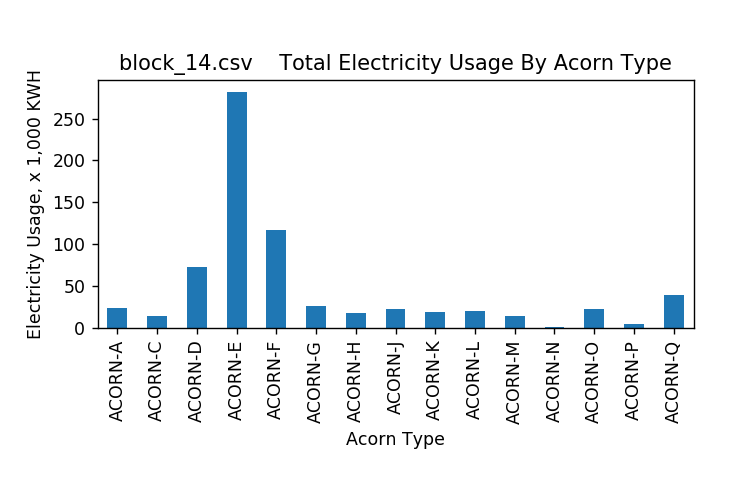


Figure 48: Total Electricity Usage by Acorn Type – Block\_14