

## Import data from text file

## Table of Contents

clear all .....	1
Setup the Import Options .....	1
take out var to use .....	3
1. Geo analysis -- The whole world .....	4
2.Histogram of success across categories(defining success using %state) .....	7
3.relationship between "spotlighted"/"staff picked" and being successful .....	8
4.relationship between goal \$\$ and being successful .....	8
5.relationship between blurb(in #length) and being successful .....	9
6.Word cloud for blurb of successful and unsuccessful projects in the largest 3 categories .....	9
7.Plot the number of projects in each category for every month of every year (a line plot with time on the x-axis). .....	10

Script for importing data from the following text file:

```
filename: C:\Users\Owner\Documents\MATLAB\Human\Kickstart\Kickstarter_2018-01\Kickstarte
```

**clear all**

```
clear all
close all
basepath = pwd
[filepath,dirname] = fileparts(basepath);
if ~strcmp(dirname,'Kickstart')
    fprintf('Your current folder is not Kickstart');
else
    addpath(basepath);
end
```

## Setup the Import Options

```
opts = delimitedTextImportOptions( "NumVariables", 23); % Specify range and delimiter % opt-
s.DataLines = [2, Inf]; opts.Delimiter = ",";
```

```
% Specify column names and types
% opts.VariableNames = ["backers_count", "blurb", "category",
    "converted_pledged_amount", "country", "created_at", "currency",
    "current_currency", "deadline", "goal", "id", "is_starrable",
    "launched_at", "name", "pledged", "slug", "source_url", "spotlight",
    "staff_pick", "state", "state_changed_at", "usd_pledged",
    "location"];
% opts.VariableTypes = ["double", "string", "categorical", "double",
    "categorical", "double", "categorical", "categorical", "double",
    "double", "double", "categorical", "double", "string", "double",
    "string", "double", "categorical", "categorical", "categorical",
    "double", "double", "string"];
% opts = setvaropts(opts, [2, 14, 16, 23], "WhitespaceRule",
    "preserve");
```

```

% opts = setvaropts(opts, 17, "zimNonNumeric", true);
% opts = setvaropts(opts, 17, "ThousandsSeparator", ",");
% opts = setvaropts(opts, [2, 3, 5, 7, 8, 12, 14, 16, 18, 19, 20, 23],
    "EmptyFieldRule", "auto");
% opts.ExtraColumnsRule = "ignore";
% opts.EmptyLineRule = "read";
%
% rand_array = randperm(100);
% rand_row_index = sort(rand_array(1:10));
Kickstarterheader
    =string({'backers_count', 'blurb', 'category', 'country', 'deadline', 'goal', 'id', 'lau
Kickstarter = [];

pre_name = "C:\Users\Owner\Documents\MATLAB\Human\Kickstart\";
month_name = "Kickstarter_2018-%02d";
file_name = "Kickstarter%03d.csv";

monthinfo = {'month'};
%headerColumn=[1 2 3 5 12 16 17 21 22 23 30 31 32 33 36];

for n = 1:12
    n
    temp_month_name = sprintf(month_name,n);

    for k=0:48
        k
        if k>0
            temp_file_name = sprintf(file_name,k);
        else
            temp_file_name = sprintf(file_name,[]);
        end

        % for i=2:length(rand_row_index)
        %     opts.DataLines=[rand_row_index(i) rand_row_index(i)]

        final_file_name = pre_name+temp_month_name+"\ "+temp_file_name;

        % Import the data
        %temp_Kickstarter = readtable(final_file_name, opts);

        [NUM,TXT,RAW]=xlsread(final_file_name);
        %todelete=[3 7 9 10 13 14 17 20 22 23 29 30 32 34 35 36 37];
        %col to delete

        headerColumn=[];
        Rawheader=string(RAW(1,:));
        for w=1:length(Kickstarterheader)

[R,headerCol]=find(cellfun(@(Rawheader)~isempty(strfind(Rawheader,Kickstarterhead
        headerColumn=[headerColumn;headerCol(1)];
        end

        rows_to_save = randperm(size(RAW,1),100);

```

```

    if any(rows_to_save(:)==1) ==1
        rows_to_save = randperm(size(RAW,1),100);
    end

    data_to_save = RAW(rows_to_save,headerColomn);

    %monthinfo=[monthinfo;n];

    Kickstarter= [Kickstarter;data_to_save];
    %z=cell2table(data_to_save)
    %z.Properties.VariableNames = data_to_save(1,:)
    % end

end

end

Kickstarter= [RAW(1,headerColomn);Kickstarter];
%Kickstarter2= [Kickstarter(2:end,:)];

```

## take out var to use

```

%ss=[];
countcate=0;
num_cat = {'num_cat'};
cate=["art","comics","crafts","dance","design","fashion","film &
    video","food","games","journalism","music","photography","publishing","technology"];
precategory2=[];
maincategory2=[];
subcategory2=[];
for h=2:size(Kickstarter2,1)

    ss=jsondecode(string(Kickstarter2(h,3)))
    %ss =[ss;jsondecode(string(data_to_save(h,3)))]
    precategory=ss.slug;
    precategory2=[precategory2;string(precategory)];

    if isempty(strfind(precategory,'/'))==1
        maincategory=precategory;
    else
        maincategory=precategory(1:strfind(precategory,'/')->1)
    end

    maincategory2=[maincategory2;string(maincategory)];
    subcategory2=[subcategory2;string(ss.name)];

    matched = 0;

    for i=1:size(cate,2)
        if strcmp(cate(i),maincategory) == 1
            num_cat = [num_cat; i];
            matched = 1;
        end
    end
end

```

```

end

%     if matched == 0
%         num_cat = [num_cat; 0];
%     end
% monthinfo=

        %string(extractAfter(data_to_save(3,13),'categories/'))
end
Kickstarter2 = [Kickstarter num_cat];

%Saved Var Mat
%load('final100.mat')

cityfullname2=[];
%city2=[];
RealCountryAbb2=[];
for k=2:size(Kickstarter2,1)
    if isnan(Kickstarter2{k,9})==1
        prelocation=[];
    else
        prelocation=jsondecode(string(Kickstarter2(k,9)))
        RealCountry=prelocation.country
        cityfullname= prelocation.short_name;

%     if isempty(strfind(precategory, '/'))==1
%         city=[];
%     else
%         city=prelocation.slug(1:strfind(prelocation.slug, '-')-1)
%     end
end
cityfullname2=[cityfullname2;string(cityfullname)];
%city2=[city2;string(city)];
RealCountryAbb2=[RealCountryAbb2,string(RealCountry)];
end

```

## 1. Geo analysis -- The whole world

```

% set colors
ocean_color = [0.7 0.8 1]; land_color = [0.9 0.9 0.8]; % map colors
ax = worldmap('World')
coastline = load('coast.mat');
plotm(coastline.lat, coastline.long)
setm(ax, 'MapProjection', 'mercator')
setm(ax, 'Origin', [0 180 0])
land = shaperead('landareas', 'UseGeoCoords', true);
geoshow(ax, land, 'FaceColor', land_color)
lakes = shaperead('worldlakes', 'UseGeoCoords', true);
geoshow(lakes, 'FaceColor', 'blue')
rivers = shaperead('worldrivers', 'UseGeoCoords', true);
geoshow(rivers, 'Color', 'blue')
cities = shaperead('worldcities', 'UseGeoCoords', true);
geoshow(cities, 'Marker', '.', 'Color', 'red')

```

```

uniquecountry= string(unique(Kickstarter2(2:end,4)));

%import countryinfo excel
opts = spreadsheetImportOptions("NumVariables", 5);
% Specify sheet and range
opts.Sheet = "Sheet1";
opts.DataRange = "A2:E249";

% Specify column names and types
opts.VariableNames =
    ["COUNTRY", "A2ISO", "A3UN", "NUMUN", "DIALINGCODE"];
opts.SelectedVariableNames =
    ["COUNTRY", "A2ISO", "A3UN", "NUMUN", "DIALINGCODE"];
opts.VariableTypes =
    ["string", "string", "string", "double", "double"];
opts = setvaropts(opts, [1, 2, 3], "WhitespaceRule", "preserve");
opts = setvaropts(opts, [1, 2, 3], "EmptyFieldRule", "auto");

% Import the data
countryinfo = readtable("C:\Users\Owner\Documents\MATLAB\Human
\Kickstart\countryinfo.xlsx", opts, "UseExcel", false);

%Clear temporary variables
clear opts
%
figure(100); clf
countryRow=[];
countryheader2=string(countryinfo{:,2});
for i= 1:length(uniquecountry)

    [Row,countryCol]=find(cellfun(@(countryheader2)~isempty(strfind(countryheader2,un
        countryRow=[countryRow;Row(1)]);

        if countryinfo{Row,1}=="Hong Kong" || countryinfo{Row,1}=="Taiwan"
            worldmap("China")
%         land = shaperead('landareas', 'UseGeoCoords', true);
%         geoshow(land, 'FaceColor', land_color)
            hold on

        elseif countryinfo{Row,1}=="Netherlands"
            worldmap("Netherlands, Kingdom of the")
%         land = shaperead('landareas', 'UseGeoCoords', true);
%         geoshow(land, 'FaceColor', land_color)
            hold on

        else
            worldmap(countryinfo{Row,1})
            hold on

        end
    land = shaperead('landareas', 'UseGeoCoords', true);
    geoshow(land, 'FaceColor', land_color)
    hold on

```

```

end

success_cate= maincategory2(Rowkick2(1:end));

successin=[];
Kickcountry=string(Kickstarter2(:,13));
[Rowkick2,Colkick2]=find(cellfun(@(Kickcountry)~isempty(strfind(Kickcountry,'succe
Kickcountryin=[Kickcountryin;Rowkick2(1)]

unsuccessin=[];
Kickcountry=string(Kickstarter2(:,13));
[Rowkick3,Colkick3]=find(cellfun(@(Kickcountry)~isempty(strfind(Kickcountry,'faile
Kickcountryin=[Kickcountryin;Rowkick3(1)]
% unsuccess_cate= maincategory2(Rowkick3(1:end));

RealCountryname1=[];
RealCountryRow=[];
uniquerealcountry=unique(RealCountryAbb2);
for i= 1:length(uniquerealcountry)

    [Row2,countryCol2]=find(cellfun(@(countryheader2)~isempty(strfind(countryheader2,
        RealCountryRow=[RealCountryRow;Row2];

    RealCountryname1=[RealCountryname1;countryinfo{Row2,1}]

end

tc = zeros(size(uniquecountry));
for q = 1:length(Kickstarter2(2:end,4))
    tc(q) = sum(Kickstarter2(2:end,4) == uniquecountry(q));
end

figure(200);clf
sgtitle('Project Country Info');
subplot(1,3,1)
bar(categorical(uniquecountry),tc)

subplot(1,3,2)
pie(tc);

subplot(1,3,3)
pie(tc,uniquecountry)

Kickcountryin=[];

Kickcountry=string(Kickstarter2(:,4));
[Rowkick1,Colkick]=find(cellfun(@(Kickcountry)~isempty(strfind(Kickcountry,'US'))),
Kickcountryin=[Kickcountryin;Rowkick1(1)]

RealCountryAbb22=RealCountryAbb2'
uniquerealcountry=uniquerealcountry'

```

```

cityinUS=cityfullname2(Rowkick1);
xxx=unique(cityinUS);
t = zeros(size(cate));
for q = 1:length(cate)
    t(q) = sum(success_cate == xx(q));
end

figure(300);clf
sgtitle('Detailed Country Info,Using Real Country');

subplot(1,3,1)
title('Country where the project is really used in')
wordcloud(RealCountryname1(:));

subplot(1,3,2)
title('Most City where the project is really coming from')
wordcloud(cityfullname2(:));

success_country=Kickstarter2(Rowkick2,4)
td = zeros(size(uniquecountry));
for q = 1:length(uniquecountry)
    td(q) = sum(success_country == uniquecountry(q));
end
subplot(1,3,3)
title('Country with most success rate')
bar(td)
%hist(td)

```

## 2.Histogram of success across categories(defining success using %state)

```

count=0;
success_cate= maincategory2(Rowkick2(1:end));
xx=unique(success_cate);
t = zeros(size(cate));
for q = 1:length(cate)
    t(q) = sum(success_cate == xx(q));
end

figure (2); clf
%for e=1:size(success_cate,1)

bar(categorical(xx), t);

xlabel('categories');
ylabel('success state');
title('success across categories');
hold on
%end

```

```
figure(100);clf
sgtitle('success across categories in percentage');

subplot(2,1,1)
labels=xx;
p=pie(t)

subplot(2,1,2)
p2=pie(t,labels)
```

### 3.relationship between "spotlighted"/"staff picked" and being successful

```
figure (3); clf
success_staff= Kickstarter2(Rowkick2,12);
success_staff_mat=cell2mat(success_staff)

m= sum(success_staff_mat == 1);
np = sum(success_staff_mat == 0);
pick=[m,np]
label2=["success/picked","success/no pick"]
for b=1:size(pick,2)
bar(categorical(label2(b)), pick(b));
hold on
end
```

### 4.relationship between goal \$\$ and being successful

```
success_goal= Kickstarter2(Rowkick2,6);
success_goal_mat=cell2mat(success_goal);

A=Kickstarter2(2:end,6)
% maxgoal=A{1,1};
% mingoal=A{1,1};
% for g=1:length(A)
%     if maxgoal<=A{g,1}
%         maxgoal=A{g,1};
%     end
%     if mingoal>=A{g,1}
%         mingoal=A{g,1};
%     end
% %mingoal=min(Kickstarter2(2:end,6))
% end
maxtotalgoal=max([A{:}]);
mintotalgoal=min([A{:}]);

[Max,Imax]=max(success_goal_mat);
[Min,Imin]=min(success_goal_mat);
```



```
x=1:length(success_goal_mat);

figure (4); clf
hist(success_goal_mat)
%plot(success_goal_mat)
```

## 5.relationship between blurb(in #length) and being successful

```
blurb_length=[];
%success_blurb_length

for r=1:length(Rowkick2)

    blurblen=length(Kickstarter2{Rowkick2(r),2});
    blurb_length=[blurb_length;blurblen];
end

%unsuccess
un_blurb_length=[];
for r=1:length(Rowkick3)

    blurblenun=length(Kickstarter2{Rowkick3(r),2});
    un_blurb_length=[un_blurb_length;blurblenun];
end

figure (5); clf
subplot(1,2,1)
hist(blurb_length)
title('sucessful blurb length')

subplot(1,2,2)
hist(un_blurb_length)
title('unsucess blurb length')
```

## 6.Word cloud for blurb of successful and unsuccessful projects in the largest 3 categories

```
% discount US figure (6); clf blurbs = erasePunctuation(Kickstart(2:end,2)); blurbs = lower(blurbs); token_blurbs = tokenizedDocument(blurbs); myStopWords = stopWords; % stopWords is a function that returns a predefined list of stop words % remove the stop words token_blurbs = removeWords(token_blurbs,myStopWords);

blurbs_bag = bagOfWords(token_blurbs); blurbs_bag = removeInfrequentWords(blurbs_bag,2);

% titles = {'world cloud for failed/canceled', 'world cloud for successful'};
% for i= 0:1
%     successful_token_blurbs = token_blurbs(Kickstarter2(Rowkick2, == i);
%     successful_blurbs_bag = bagOfWords(successful_token_blurbs);
```

```
%      subplot(1,4,i+2);
%      wordcloud(successful_blurbs_bag);
%      title(titles{i+1});
% end
```

## 7. Plot the number of projects in each category for every month of every year (a line plot with time on the x-axis).

```
months = zeros(size(Kickstarter2,1),1); month_1_catsumf=[];

ttt = [];
for i = 1:4900:58800

    %months(49*length(rows_to_save)*(i-1)+1:49*length(rows_to_save)*i)=i;
    xo=unique(maincategory2);
    tt = zeros(size(i,15));
    for q = 1:length(cate)
        tt(q) = sum(maincategory2(i:(4900 + i -1)) == xo(q));
    end
    ttt = [ttt; tt];
    %      month_category =Kickstarter2(:,16) == category);
    %      success_categories_category = success_categories(maincategory2
    %      == category);
    % end
end
figure (7); clf
plot(1:12,ttt)
```

*Published with MATLAB® R2019a*