Merit America Google Data Analytics Certificate Capstone Project

Case Study: 2

Bellabeat Analysis

Author:Braxton Greer

Date: 11/06/2023

Email: braxtonkgreer@gmail.com

GitHub

Linkedin

Ask

This is the Capstone Project for the Google Data Analytics Certificate through Merit America. The task given was to complete a case study as if I were a Marketing Analyst for the company, BellaBeat, a tech manufacturer of health products focused for women.

The Company

The company, BellaBeat, sells and manufactures health-focused products aimed at a female audience. These products collect data on activities, stress, sleep, and reproductive health, allowing BellaBeat to give women power with knowledge about their health.

Stakeholders

Urska Srsen: Co-Founder and Chief Creative Officer

Sando Mur: Co-Founder and Mathematician

Question

What are some trends in smart device usage?

How could these trends apply to BellaBeat's Customers?

How could these trends influence Bellabeat's marketing?

Deliverables

- A clear summary
- A description of all data sources
- Documentation of any cleaning or manipulation of data
- A summary of the analysis
- Supporting visualizations
- Recommendations based on analysis

Prepare

Data Source

FitBit Fitness Tracker Data

- 18 Datasets were generated by to respondents to a survey via Amazon Mechanical Turk between 04/12/2013-05/12/2016
- 33 eligible FitBit Users consented to the submission of the tracker data

Limitations

The data collected used 30 users during a period of a month, in the year 2016. The data is outdated, and the sample size is considerably small. There were no descriptors or information on the participants besides the data collected from the FitBit device leaving the possibility for an

outside source from the targeted demographic. However, for the project at hand which is showcasing my ability to analyze this data will work.

Process

Clean data in Excel

I chose to work with the following datasets in an Excel Workbook:

- dailyActivity_merged
- sleepDay_merged

My decision to work with these data sets out of the 18 possible, was because most of the data could be found in the “dailyActivity_merged” dataset and/or not relevant.

Cleaning Steps

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Id	ActivityDate	TotalSteps	TotalDistanc	TrackerDista	LoggedActiv	VeryActiveD	ModeratelyA	LightActiveI	SedentaryA	VeryActiveM	FairlyActive	LightlyActiv	SedentaryM	Calories
2	1503960366	4/12/2016	13162	8.5	8.5	0	1.879999995	0.550000119	6.059999943	0	25	13	328	728	1985
3	1503960366	4/13/2016	10735	6.96999979	6.96999979	0	1.570000052	0.689999976	4.710000038	0	21	19	217	776	1797
4	1503960366	4/14/2016	10460	6.739999771	6.739999771	0	2.440000057	0.400000006	3.910000086	0	30	11	181	1218	1776
5	1503960366	4/15/2016	9762	6.28000021	6.28000021	0	2.140000105	1.25999999	2.829999924	0	29	34	209	726	1745
6	1503960366	4/16/2016	12669	8.159999847	8.159999847	0	2.710000038	0.409999964	5.039999962	0	36	10	221	773	1863
7	1503960366	4/17/2016	9705	6.480000019	6.480000019	0	3.190000057	0.7799999714	2.50999999	0	38	20	164	539	1728
8	1503960366	4/18/2016	13019	8.590000153	8.590000153	0	3.25	0.6399999857	4.710000038	0	42	16	233	1149	1921
9	1503960366	4/19/2016	15506	9.880000114	9.880000114	0	3.529999971	1.320000052	5.03000021	0	50	31	264	775	2035
10	1503960366	4/20/2016	10544	6.679999828	6.679999828	0	1.960000038	0.4799999893	4.239999771	0	28	12	205	818	1786
11	1503960366	4/21/2016	9819	6.340000153	6.340000153	0	1.340000033	0.349999994	4.650000095	0	19	8	211	838	1775
12	1503960366	4/22/2016	12764	8.130000114	8.130000114	0	4.760000229	1.120000005	2.24000001	0	66	27	130	1217	1827
13	1503960366	4/23/2016	14371	9.039999962	9.039999962	0	2.809999943	0.8700000048	5.360000134	0	41	21	262	732	1949
14	1503960366	4/24/2016	10039	6.409999847	6.409999847	0	2.920000076	0.2099999934	3.279999971	0	39	5	238	709	1788
15	1503960366	4/25/2016	15355	9.800000191	9.800000191	0	5.289999962	0.5699999928	3.940000057	0	73	14	216	814	2013
16	1503960366	4/26/2016	13755	8.789999962	8.789999962	0	2.329999924	0.9200000167	5.539999962	0	31	23	279	833	1970
17	1503960366	4/27/2016	18134	12.21000004	12.21000004	0	6.400000095	0.4099999964	5.409999847	0	78	11	243	1108	2159
18	1503960366	4/28/2016	13154	8.529999733	8.529999733	0	3.539999962	1.159999967	3.789999962	0	48	28	189	782	1898
19	1503960366	4/29/2016	11181	7.150000095	7.150000095	0	1.059999943	0.5	5.579999924	0	16	12	243	815	1837
20	1503960366	4/30/2016	14673	9.25	9.25	0	3.559999943	1.419999957	4.269999981	0	52	34	217	712	1947
21	1503960366	5/1/2016	10602	6.809999943	6.809999943	0	2.289999962	1.600000024	2.920000076	0	33	35	246	730	1820
22	1503960366	5/2/2016	14727	9.710000038	9.710000038	0	3.210000038	0.5699999928	5.920000076	0	41	15	277	798	2004
23	1503960366	5/3/2016	15103	9.659999847	9.659999847	0	3.730000019	1.049999952	4.880000114	0	50	24	254	816	1990
24	1503960366	5/4/2016	11100	7.150000095	7.150000095	0	2.460000038	0.8700000048	3.819999933	0	36	22	203	1179	1819
25	1503960366	5/5/2016	14070	8.899999619	8.899999619	0	2.920000076	1.080000043	4.880000114	0	45	24	250	857	1959
26	1503960366	5/6/2016	12159	8.029999733	8.029999733	0	1.970000029	0.25	5.809999943	0	24	6	289	754	1896
27	1503960366	5/7/2016	11992	7.710000038	7.710000038	0	2.460000038	2.119999886	3.130000114	0	37	46	175	833	1821
28	1503960366	5/8/2016	10060	6.579999924	6.579999924	0	3.529999971	0.3199999928	2.730000019	0	44	8	203	574	1740
29	1503960366	5/9/2016	12022	7.71999979	7.71999979	0	3.450000048	0.5299999714	3.74000001	0	46	11	206	835	1819
30	1503960366	5/10/2016	12207	7.769999981	7.769999981	0	3.349999905	1.159999967	3.25999999	0	46	31	214	746	1859
31	1503960366	5/11/2016	12770	8.130000114	8.130000114	0	2.559999943	1.00999999	4.550000191	0	36	23	251	669	1783
32	1503960366	5/12/2016	0	0	0	0	0	0	0	0	0	0	0	1440	0
33	1624580081	4/12/2016	8163	5.309999943	5.309999943	0	0	0	5.309999943	0	0	0	146	1294	1432

I went through each data set and utilizing filters removed any duplicate information. I also changed all of the date formatting from mm/dd/yyyy hr:min:sec to just mm/dd/yyyy.

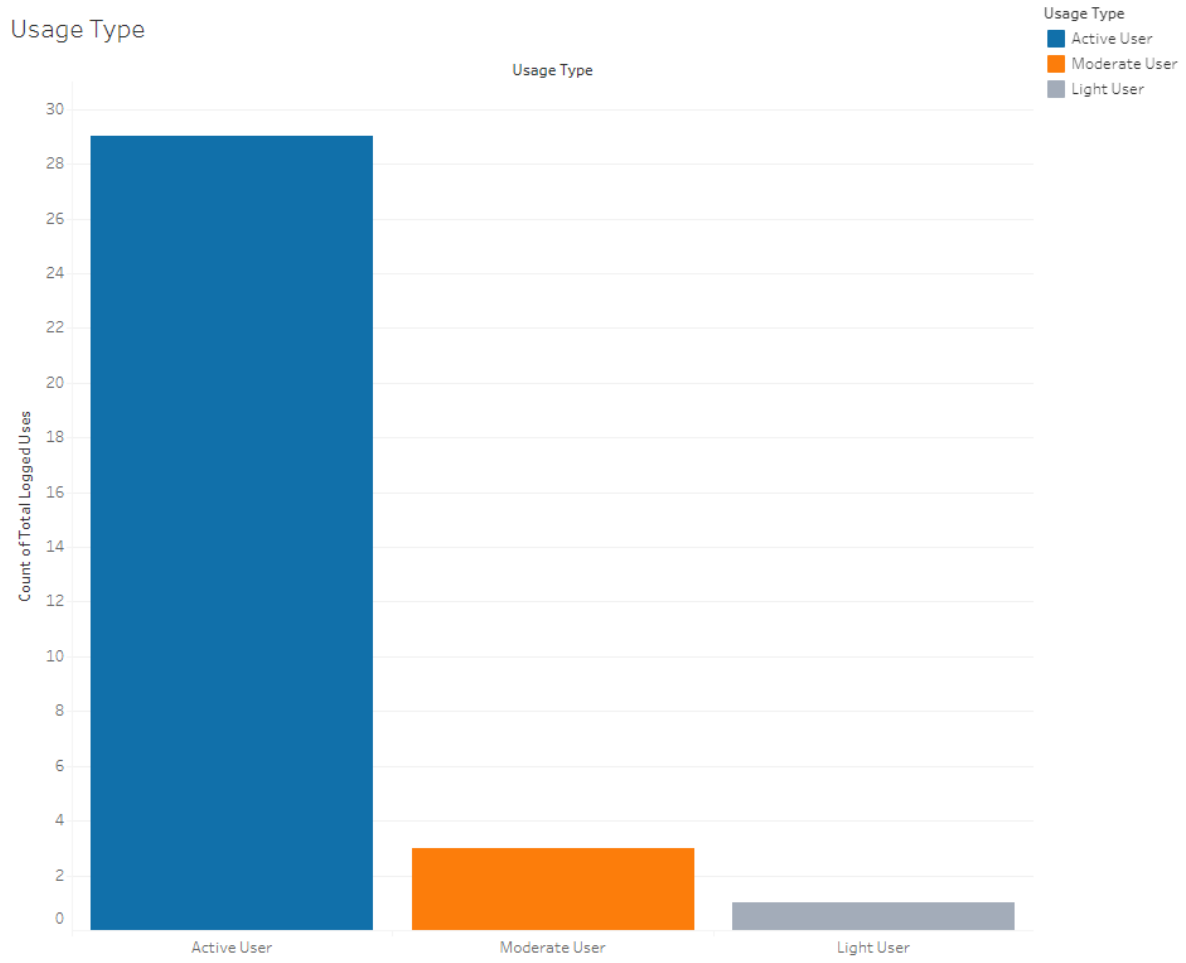
Analyze

All datasets were uploaded to BigQuery under the project name “bellabeats-casestudy-capstone”.

To start I wanted to divide users into subgroups for how often the tracker was worn within the time period.

- Active user: 21-31 days
- Moderate user:11-20 days
- Light users:0-10 days

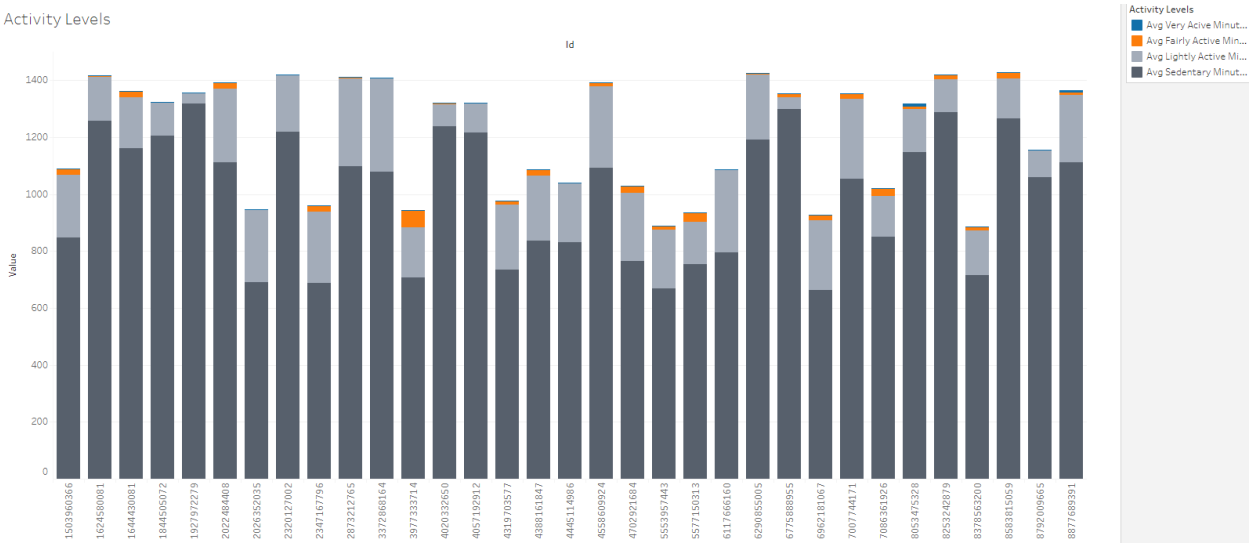
```sql
SELECT Id, count(Id) AS Total_Logged_Uses,
CASE
  When count (Id) BETWEEN 21 And 31 Then 'Active User'
  When count (Id) Between 11 and 20 then 'Moderate User'
  When count (Id) Between 0 and 10 Then 'Light User'
End as Usage_Type
FROM `bellabeats-casestudy-capstone.fitbit.DailyActivity`
group by Id
order by Total_Logged_Uses desc;
```



Next was to look at the minimum, maximum, and average of total steps, total distance, calories, and activity levels grouped by ID.

```sql
SELECT
  Id,
  Min(TotalSteps) As Min_Total_Steps,
  Max(TotalSteps) As Max_Total_Steps,
  Avg (TotalSteps) As Avg_Total_Steps,
  Min (TotalDistance) As Min_Total_Distance,
  Max (TotalDistance) As Min_Total_Distance,
  Avg (TotalDistance) As Avg_Total_Distance,
  Min (Calories) As Min_Calories,
  Max (Calories) as Max_Calories,
  Avg (Calories) as Avg_Calories,
  Min(VeryActiveMinutes) As Min_Very_Avtive_Minutes,
  Max (VeryActiveMinutes) as Max_Very_Active_Minutes,
  Avg (VeryActiveDistance) as Avg_Very_Acive_Minutes,
  Min (FairlyActiveMinutes) as Min_Fairly_Active_Minutes,
  Max (FairlyActiveMinutes) as Max_Fairly_Active_Minutes,
  Avg (FairlyActiveMinutes) as Avg_Fairly_Active_Minutes,
  Min (LightlyActiveMinutes) as Min_Lightly_Active_Minutes,
  Max (LightlyActiveMinutes) as Max_Lightly_Active_Minutes,
  Avg (LightlyActiveMinutes) as Avg_Lightly_Active_Minutes,
  Min (SedentaryMinutes) as Min_Sedentary_Minutes,
  Max (SedentaryMinutes) as Max_Sedentary_Minutes,
  Avg (SedentaryMinutes) as Avg_Sedentary_Minutes
FROM `bellabeats-casestudy-capstone.fitbit.DailyActivity`
Group BY Id;
```

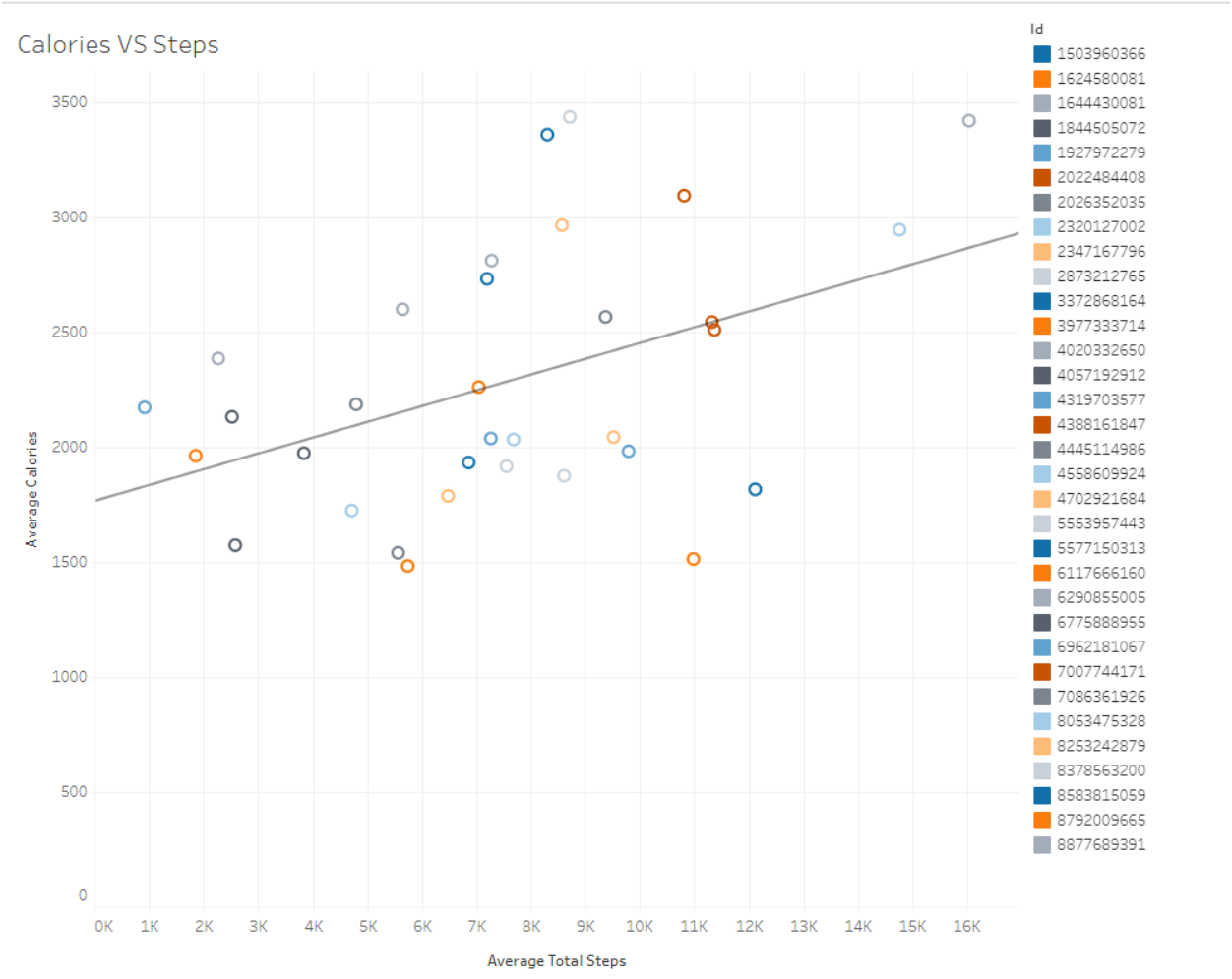
And to use the average of the minutes to narrow down the results.



This showed that most users logged a sedentary activity level a majority of the time.

The next set of data I wanted to look at was to see if there was a correlation between the number of calories burned by the average amount of steps taken.

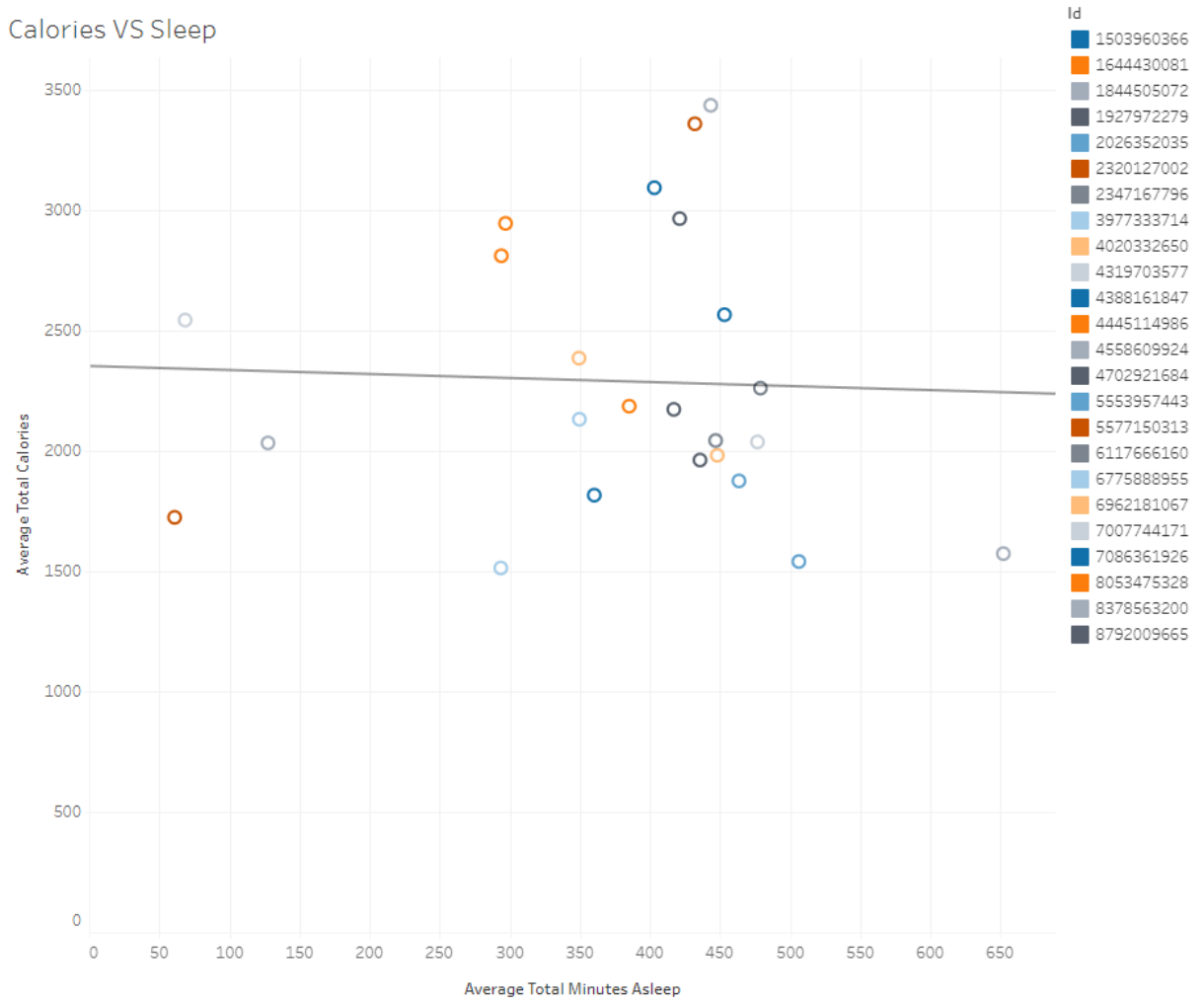
```sql
SELECT
  Id,
  Avg (TotalSteps) as Avg_Total_Steps,
  Avg (Calories) as Avg_Calories
FROM `bellabeats-casestudy-capstone.fitbit.DailyActivity`
Group BY
  Id;
```



There was a clear correlation between more steps taken and more calories burned.

Lastly, I wanted to check on the correlation between the calories burned by the average minutes slept.

```sql
select
  Activity.Id,
  Avg(Activity.Calories) As Avg_Total_Calories,
  Avg(Sleep.TotalMinutesAsleep) as Avg_Total_Minutes_Asleep
From
  `bellabeats-casestudy-capstone.fitbit.DailyActivity` as Activity
JOIN
  `bellabeats-casestudy-capstone.fitbit.DailySleep` as Sleep ON Activity.Id = Sleep.Id
GROUP BY Activity.Id;
```



This data did not show any noteworthy correlation, if any at all.

Share

The link below is for my Tableau Project

https://public.tableau.com/app/profile/braxton.greer/viz/BellaBeatsCapstone_17010604960810/Dashboard1

Act

Trends

- Users spent most of their time in the sedentary category.
- A majority of the users were active users, followed by moderate users.
- The more steps taken equated to more calories burned.

Recommendations

- Incentives for achieving movement goals such as badges to show off to the community and/or see accomplishments. Points can also be acquired and used for redeeming rewards.
- Incentives for consistently using devices such as daily awards and days' work streaks which allows the user to see their achievements.