

---

# **Car Brand - User Generated Content Data Analysis**

# Agenda

---

- Text-Mining with R studio
- Brand Strength Evaluation
- Find insight through Regression Analysis
- Strategy and Recommendation

---



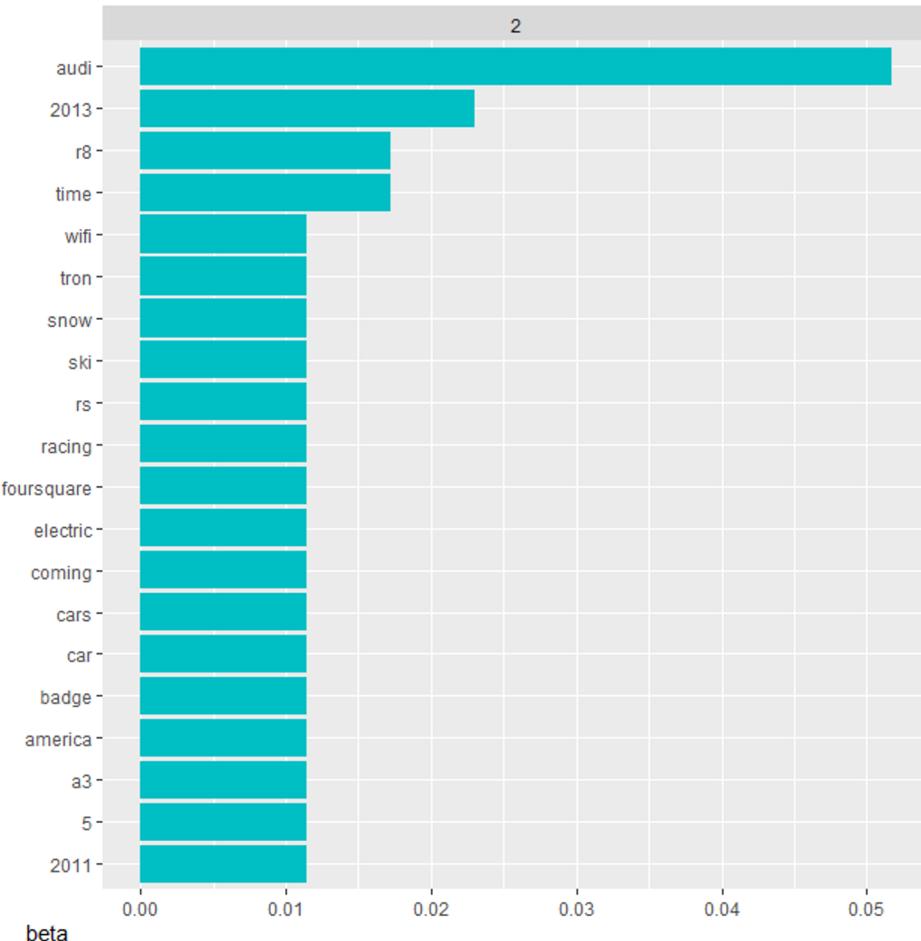
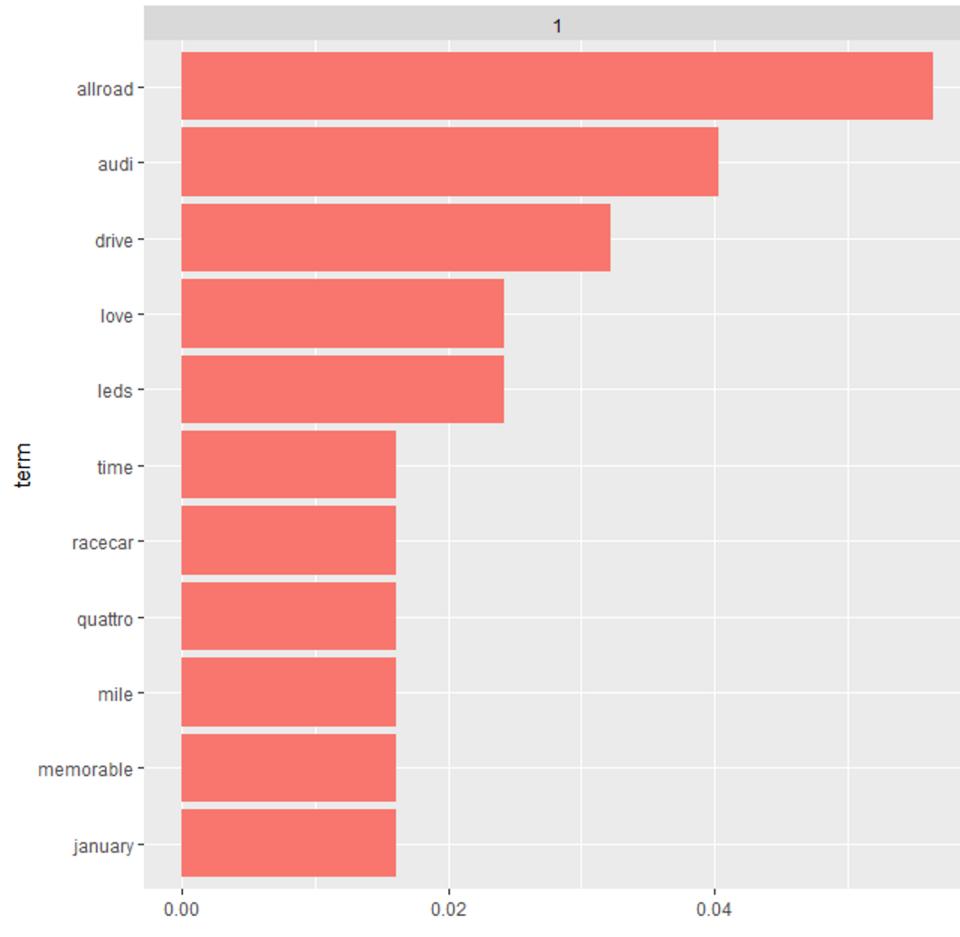
# *Text mining with R studio*

# Calculating the scores

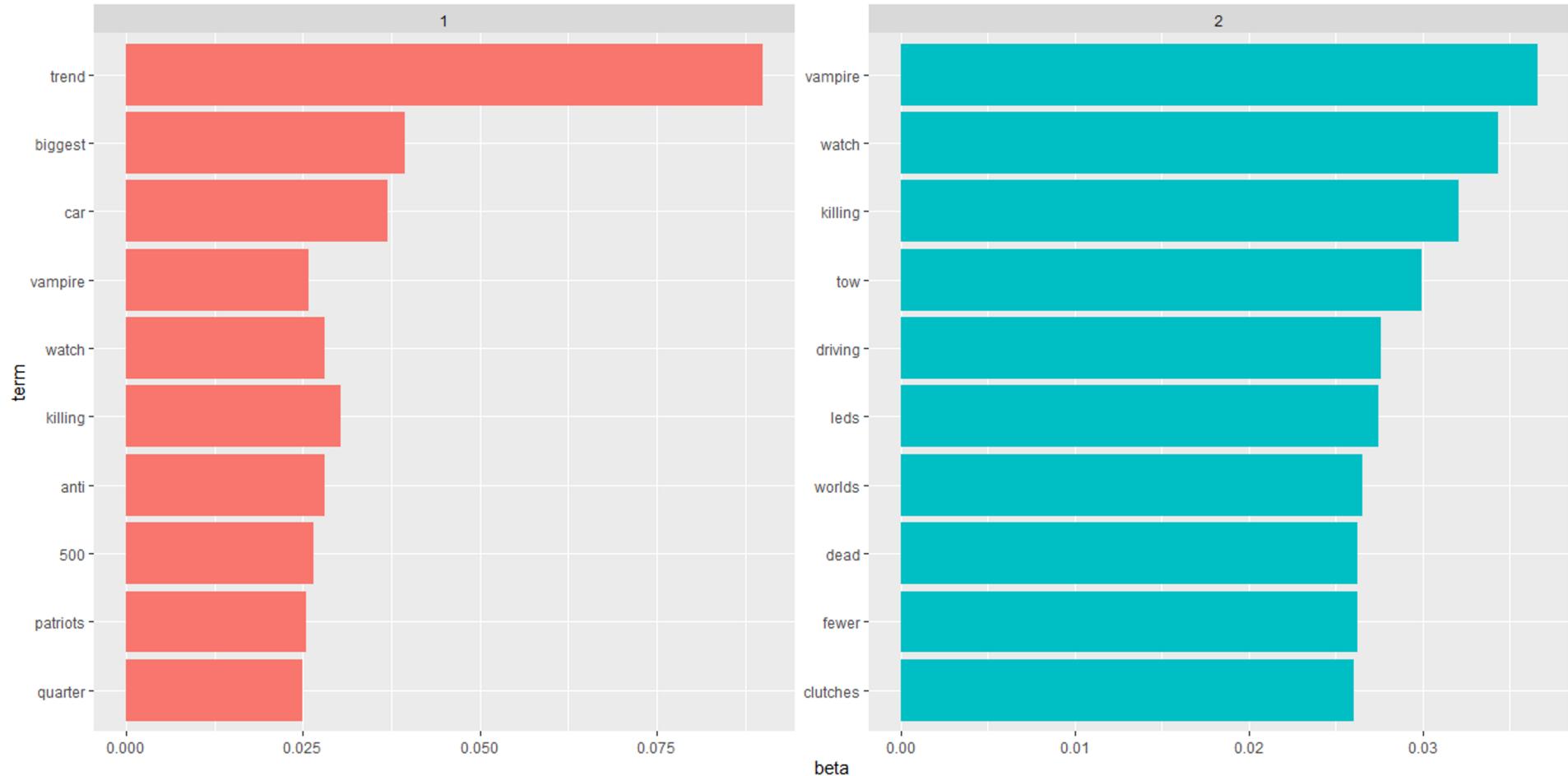
	A	B	C	E	F	G	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	id	company_url	shares	C2B_likes	total_com	message		afinnscore	bingscore	anger	anticipatio	disgust	fear	joy	negative	positive	sadness	surprise	trust
2	1	1 Acura	0	30	17	driving to work may not be one of	3	1	0	0	0	0	1	0	1	0	0	0	0
3	2	2 Acura	0	55	53	good morning acura fans! what are NA	1	0	0	0	0	0	0	0	0	1	0	0	0
4	3	3 Acura	0	53	31	remember to set your clocks ahead	-3	-1	1	0	0	0	0	0	1	1	1	0	0
5	4	4 Acura	19	180	60	excited for the new acura ilx? get c	5	1	0	1	0	0	1	0	3	0	3	0	1
6	5	5 Acura	2	87	15	you may remember tyson hugie as	7	2	0	3	0	0	0	3	0	3	0	1	2
7	6	6 Acura	0	24	9	it's quick and easy to retrieve you	1	2	0	0	0	0	0	0	0	1	0	0	0
8	7	7 Acura	0	30	60	happy monday! what did you and	3	1	0	1	0	0	0	1	0	1	0	0	1
9	8	8 Acura	0	320	41	like this status if your acura naviga	-3	0	0	0	0	0	0	0	1	1	1	0	1
10	9	9 Acura	0	31	18	are you lucky enough to steal away	1	0	1	1	0	1	2	1	2	1	1	1	0
11	10	10 Acura	0	61	6	congratulations to the 5 acura fans	5	5	1	1	1	1	2	1	3	1	1	1	1
12	11	11 Acura	0	40	26	do you have a story to share about	4	1	0	1	0	0	0	1	0	2	0	0	1
13	12	12 Acura	0	27	111	what song really gets you in the m	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14	13	13 Acura	0	40	5	happy leap day everyone! get out	5	2	0	3	0	0	2	0	3	0	0	0	2
15	14	14 Acura	0	323	29	like this status if you're already da	NA	0	0	0	0	0	0	0	1	0	0	0	0
16	15	15 Acura	0	43	71	do you have any accessories on yo	NA	0	0	0	1	0	0	0	0	0	0	0	0
17	16	16 Acura	0	91	20	if you're out and about this weeke	NA	0	1	0	0	0	0	0	1	0	0	0	0
18	17	17 Acura	0	126	184	can you guess the year and model	NA	0	0	0	0	0	0	0	1	0	0	1	0
19	18	18 Acura	0	400	656	if you could wake up tomorrow wi	NA	0	1	0	0	0	0	0	0	0	0	0	0
20	19	19 Acura	0	41	20	we love it when our fans share pic	6	3	0	1	0	0	3	0	3	0	0	0	2
21	20	20 Acura	0	44	68	driving in the snow can be tricky if	NA	-1	0	0	0	0	0	0	1	0	0	0	1
22	21	21 Acura	0	67	3	happy president's day! the preside	4	1	1	2	0	0	1	1	3	1	0	0	1
23	22	22 Acura	0	28	2	it's quick and easy to retrieve you	1	2	0	0	0	0	0	0	1	0	0	0	0
24	23	23 Acura	0	48	94	if you could give away an acura, w	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
25	24	24 Acura	0	48	4	we're celebrating president's day	5	NA	0	1	0	0	1	0	2	0	0	0	0
26	25	25 Acura	0	30	12	the glamourai knows fashion like t	5	4	0	2	0	0	2	0	3	0	1	1	1
27	26	26 Acura	0	58	15	congratulations to elliot l. scheiner	7	3	0	3	0	0	2	2	4	0	1	3	0
28	27	27 Acura	0	43	7	last week we auctioned off a custo	5	2	0	3	0	0	2	0	2	0	0	0	0
29	28	28 Acura	0	48	20	happy valentine's day acura fans!	3	2	0	1	0	0	1	0	1	0	0	0	1
30	29	29 Acura	0	68	40	love is in the air! has your acura be	3	1	0	0	0	0	1	0	1	0	0	0	0
31	30	30 Acura	0	258	70	the future is here. check out our n	2	1	0	0	0	0	0	0	1	0	0	0	1
32	31	31 Acura	0	70	24	last year tyson hugie reached on i	4	1	0	0	0	0	0	0	2	0	0	0	0

complete results

# Top Terms for Audi (Positive Afinnscore, seed = 100)



# Top Terms for Audi (Negative Afinnscore, seed = $10^7$ )



---



# *Brand Strength Evaluation*

# Data Preprocess

---

Raw CSV file provided by professor.



Processed CSV file with sentimental analysis variables.



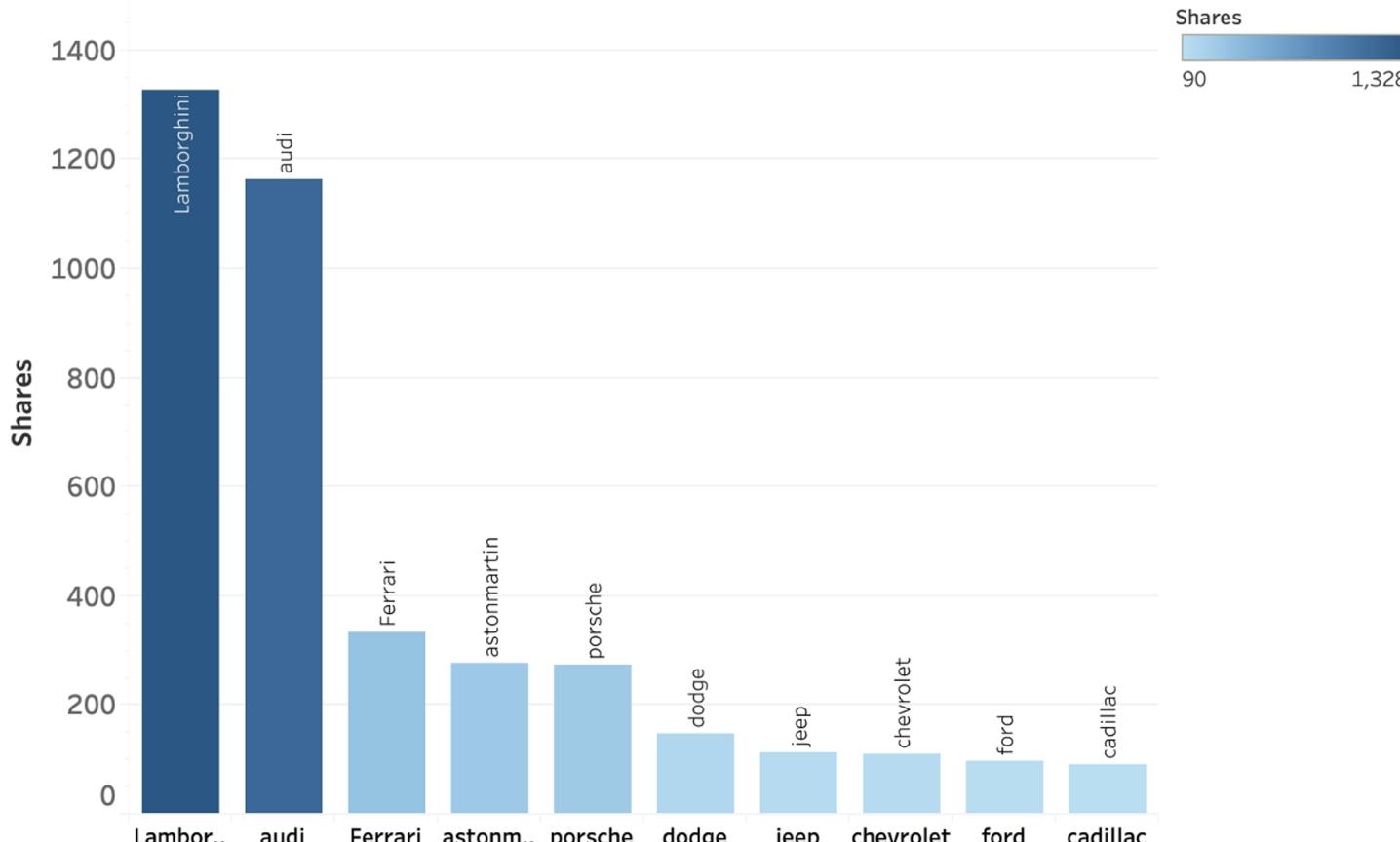
CSV file data grouped by the company name.

# Data Set For Analysis

---

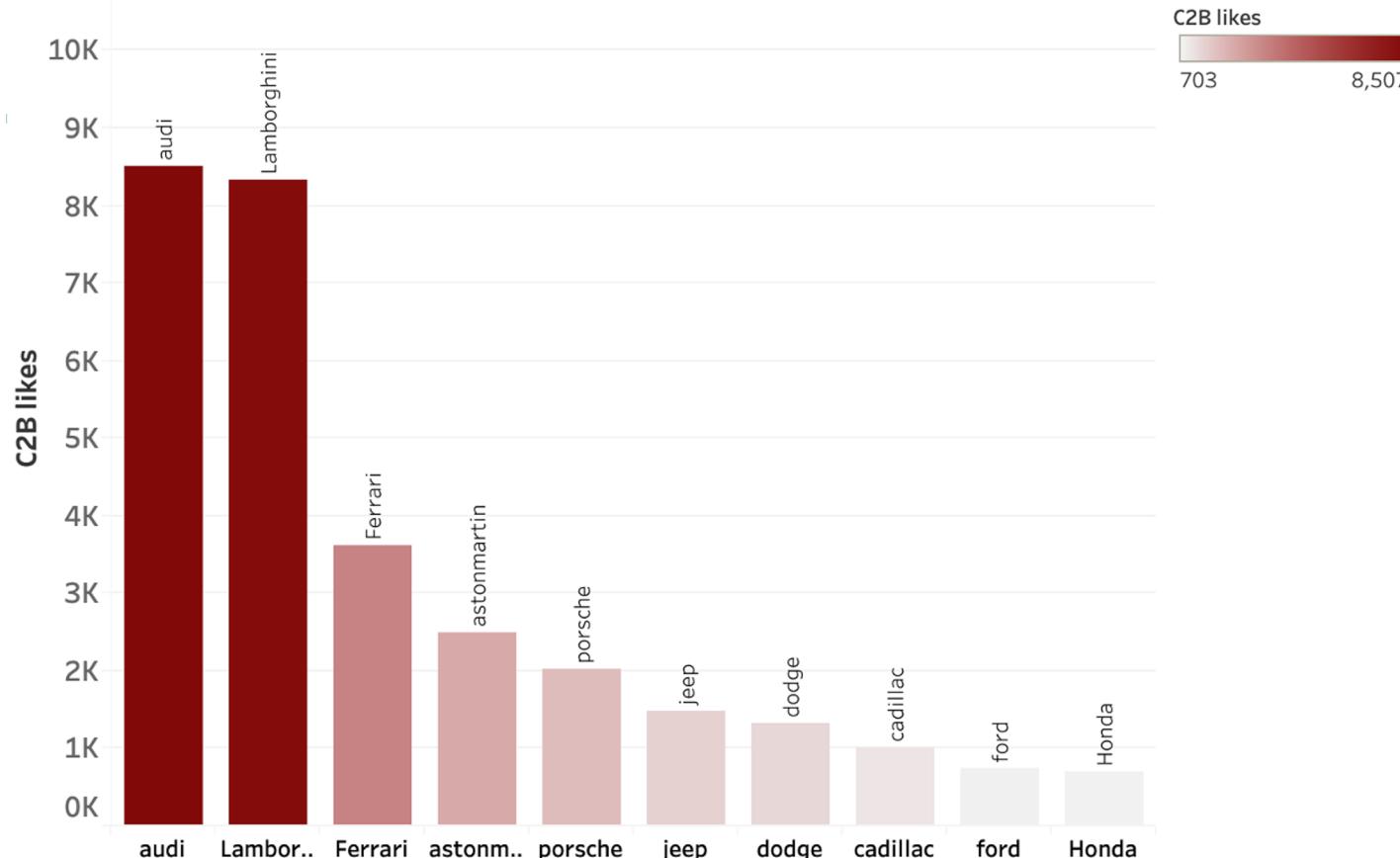
- **3 Dependent Variables :**  
= Shares + Likes + Comments
- **24 Independent Variables :**  
= 10 Emotion Related + 5 Topic Related + 3 Time Related + 6 Type Related

# Competitiveness\_Analysis - Shares



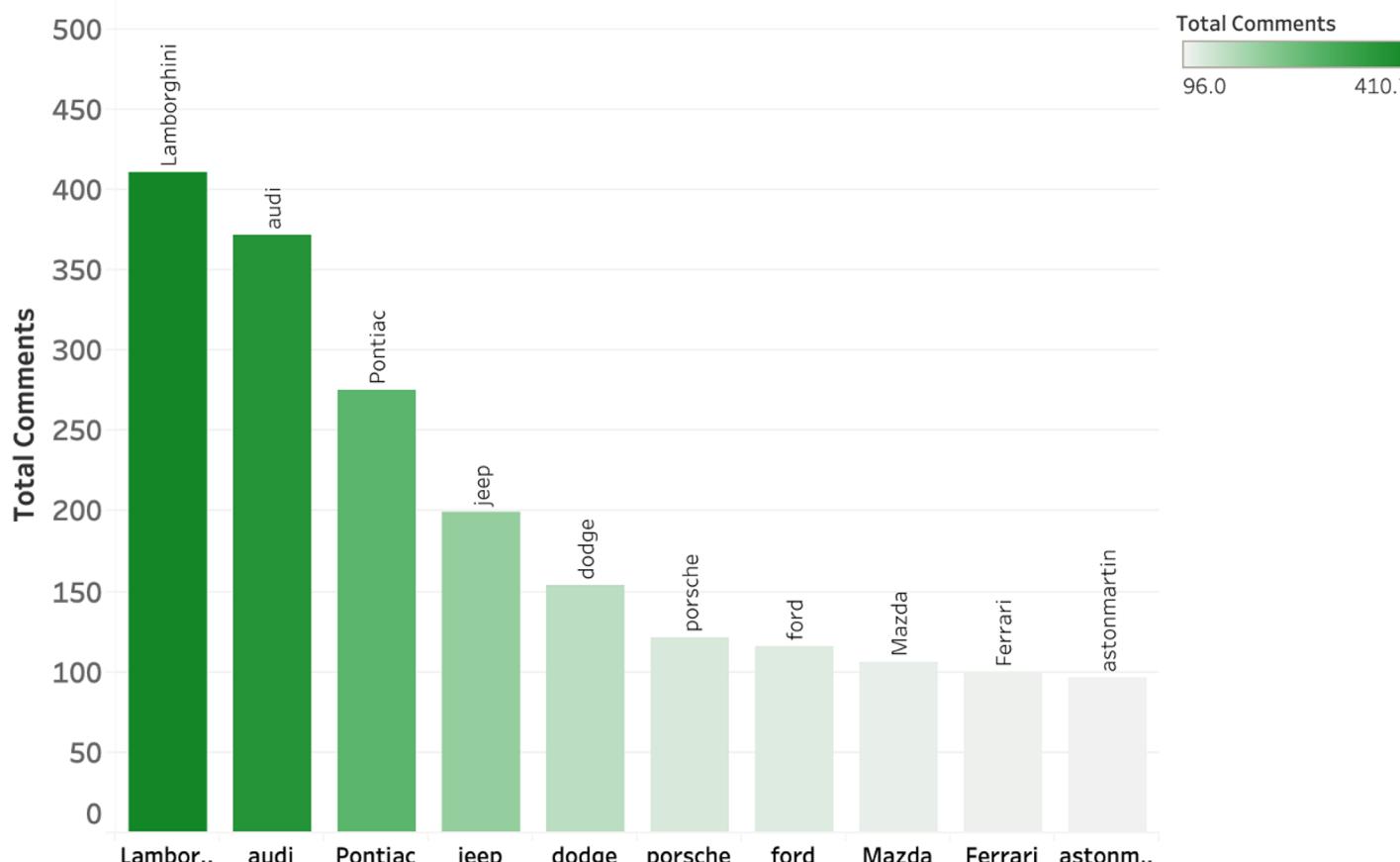
Sum of Shares for each Company Urlname. Color shows sum of Shares. The marks are labeled by Company Urlname. The view is filtered on Company Urlname, which keeps 10 of 43 members.

# Competitiveness\_Analysis - C2B likes



Sum of C2B likes for each Company Urlname. Color shows sum of C2B likes. The marks are labeled by Company Urlname. The view is filtered on Company Urlname, which keeps 10 of 43 members.

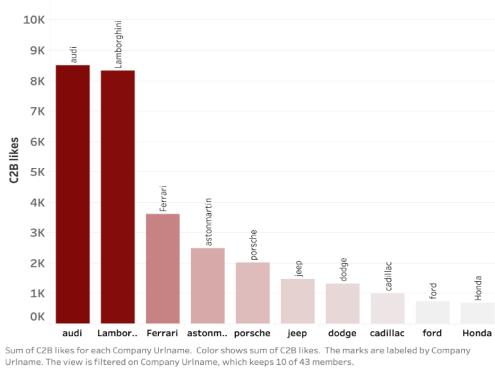
# Competitiveness\_Analysis - Total Comments



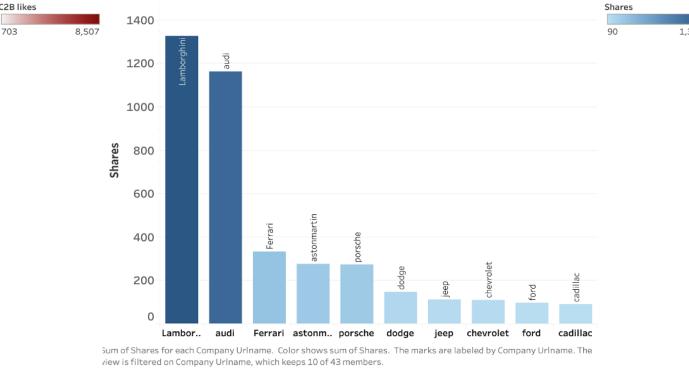
Sum of Total Comments for each Company Urlname. Color shows sum of Total Comments. The marks are labeled by Company Urlname. The view is filtered on Company Urlname, which keeps 10 of 43 members.

# Competitiveness\_Analysis - Likes, Shares, Comments

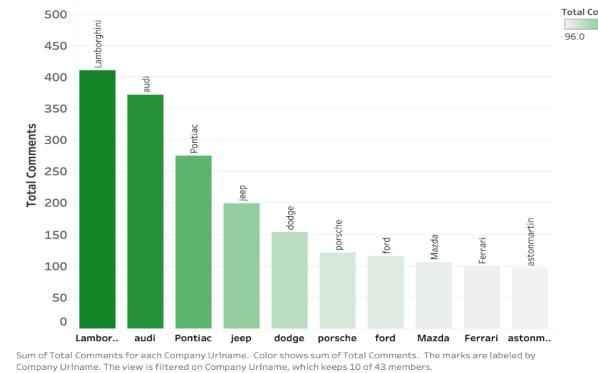
 703 - 8,507



Top 10  
C2B\_Likes Car Brand



Top 10  
Shares Car Brand



Top 10  
Comments Car Brand

***First Tier : Lamborghini, Audi, Astonmartin, Porsche , Ferrari***

# Regression\_Model ( Emotion v.s Shares, Likes, Comments )

---

## Likes

Call:  
lm(formula = C2B\_likes ~ anticipation + disgust + negative + trust)

Residuals:

Min	1Q	Median	3Q	Max
-2162.5	-806.3	-419.3	430.8	6507.8

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	641.3	914.9	0.701	0.48759
anticipation	1862.3	678.9	2.743	0.00924 **
disgust	6770.2	2163.7	3.129	0.00336 **
negative	-2640.1	1000.9	-2.638	0.01203 *
trust	-1371.2	682.6	-2.009	0.05169 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1618 on 38 degrees of freedom  
Multiple R-squared: 0.2802, Adjusted R-squared: 0.2044  
F-statistic: 3.698 on 4 and 38 DF, p-value: 0.01225

## Comments

Call:  
lm(formula = total\_comments ~ anticipation + disgust + negative + surprise + trust)

Residuals:

Min	1Q	Median	3Q	Max
-100.26	-47.82	-19.49	19.95	315.00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	81.73	47.78	1.711	0.09555 .
anticipation	72.48	40.33	1.797	0.08050 .
disgust	312.99	111.61	2.804	0.00799 **
negative	-104.92	51.65	-2.031	0.04944 *
surprise	-58.44	63.02	-0.927	0.35976
trust	-45.49	35.26	-1.290	0.20496

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 83.47 on 37 degrees of freedom  
Multiple R-squared: 0.2218, Adjusted R-squared: 0.1167  
F-statistic: 2.109 on 5 and 37 DF, p-value: 0.08602

## Shares

Call:  
lm(formula = shares ~ anticipation + disgust + negative + trust)

Residuals:

Min	1Q	Median	3Q	Max
-292.60	-110.83	-55.22	33.38	1071.75

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	89.46	136.16	0.657	0.51511
anticipation	266.12	101.04	2.634	0.01215 *
disgust	881.32	322.03	2.737	0.00938 **
negative	-389.03	148.97	-2.612	0.01283 *
trust	-195.10	101.59	-1.921	0.06232 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 240.9 on 38 degrees of freedom  
Multiple R-squared: 0.2509, Adjusted R-squared: 0.172  
F-statistic: 3.181 on 4 and 38 DF, p-value: 0.02385

Anticipation

Trust

Disgust

Negative

# Regression\_Model (**Topic** v.s Shares, Likes, Comments)

---

## Likes

```
Call:  
lm(formula = C2B_likes ~ adventure + luxury + affordable)
```

### Residuals:

Min	1Q	Median	3Q	Max
-2296.6	-651.9	-97.4	371.1	5919.2

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-209.2	354.2	-0.591	0.5581
adventure	1741.0	588.0	2.961	0.0052 **
luxury	1056.3	580.5	1.820	0.0765 .
affordable	1039.0	554.9	1.872	0.0686 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1502 on 39 degrees of freedom

Multiple R-squared: 0.3639, Adjusted R-squared: 0.315

F-statistic: 7.438 on 3 and 39 DF, p-value: 0.0004699

## Comments

```
Call:  
lm(formula = total_comments ~ safety + adventure + affordable)
```

### Residuals:

Min	1Q	Median	3Q	Max
-117.140	-43.854	-8.367	24.683	250.362

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	35.17	19.08	1.843	0.07293 .
safety	17.70	28.23	0.627	0.53431
adventure	85.87	28.46	3.017	0.00448 **
affordable	30.30	30.12	1.006	0.32057

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 82.3 on 39 degrees of freedom

Multiple R-squared: 0.2027, Adjusted R-squared: 0.1413

F-statistic: 3.304 on 3 and 39 DF, p-value: 0.03007

## Shares

```
Call:  
lm(formula = shares ~ adventure + luxury + affordable)
```

### Residuals:

Min	1Q	Median	3Q	Max
-314.63	-92.29	-42.21	62.22	826.67

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-51.82	52.82	-0.981	0.33258
adventure	240.06	87.69	2.738	0.00928 **
luxury	148.59	86.57	1.716	0.09402 .
affordable	168.84	82.76	2.040	0.04815 *

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 224 on 39 degrees of freedom

Multiple R-squared: 0.3353, Adjusted R-squared: 0.2841

F-statistic: 6.557 on 3 and 39 DF, p-value: 0.001071

Adventure

Affordable

Luxury

# Regression\_Model ( Time v.s Shares, Likes, Comments )

---

## Likes

```
Call:  
lm(formula = C2B_likes ~ afternoon + morning + night)
```

### Residuals:

	Min	1Q	Median	3Q	Max
	-2698	-983	-584	-179	101540

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	653.06	185.05	3.529	0.000423 ***
afternoon	688.98	202.77	3.398	0.000687 ***
morning	2045.20	293.49	6.969	3.86e-12 ***
night	65.35	204.59	0.319	0.749422

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3173 on 3271 degrees of freedom  
Multiple R-squared: 0.02465, Adjusted R-squared: 0.02376  
F-statistic: 27.56 on 3 and 3271 DF, p-value: < 2.2e-16

## Comments

```
Call:  
lm(formula = total_comments ~ afternoon + morning + night)
```

### Residuals:

	Min	1Q	Median	3Q	Max
	-102.7	-65.1	-47.7	-8.8	5334.2

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	66.095	12.268	5.388	7.64e-08 ***
afternoon	27.695	13.443	2.060	0.0395 *
morning	36.570	19.457	1.879	0.0603 .
night	-2.394	13.564	-0.177	0.8599

---

Signif. codes: 0 ‘\*\*\*\*’ 0.001 ‘\*\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 210.4 on 3271 degrees of freedom  
Multiple R-squared: 0.005401, Adjusted R-squared: 0.004489  
F-statistic: 5.921 on 3 and 3271 DF, p-value: 0.0005029

## Shares

```
Call:  
lm(formula = shares ~ afternoon + morning + night)
```

### Residuals:

	Min	1Q	Median	3Q	Max
	-398.6	-127.6	-80.5	-35.5	6011.4

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	78.252	24.422	3.204	0.00137 **
afternoon	70.851	26.760	2.648	0.00814 **
morning	320.315	38.733	8.270	< 2e-16 ***
night	4.231	27.001	0.157	0.87550

---

Signif. codes: 0 ‘\*\*\*\*’ 0.001 ‘\*\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 418.7 on 3271 degrees of freedom  
Multiple R-squared: 0.03092, Adjusted R-squared: 0.03003  
F-statistic: 34.79 on 3 and 3271 DF, p-value: < 2.2e-16

Morning

Afternoon

# Regression\_Model ( Post type v.s Shares, Likes, Comments )

## Likes

Call:  
lm(formula = C2B\_likes ~ link + music + photo + question + swf + video)

Residuals:

Min	1Q	Median	3Q	Max
-2051	-1065	-347	-17	100830

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	328.7	148.9	2.207	0.0273 *
link	172.1	176.2	0.977	0.3287
music	1762.6	1811.5	0.973	0.3306
photo	1723.7	174.9	9.853	< 2e-16 ***
question	-328.7	479.8	-0.685	0.4933
swf	133.0	368.6	0.361	0.7182
video	838.8	211.4	3.967	7.42e-05 ***

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3127 on 3268 degrees of freedom  
Multiple R-squared: 0.05354, Adjusted R-squared: 0.0518  
F-statistic: 30.81 on 6 and 3268 DF, p-value: < 2.2e-16

## Comments

Call:  
lm(formula = total\_comments ~ link + music + photo + question + swf + video)

Residuals:

Min	1Q	Median	3Q	Max
-130.3	-75.3	-25.0	3.0	5297.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	100.506	9.828	10.226	< 2e-16 ***
link	-69.493	11.627	-5.977	2.52e-09 ***
music	-23.506	119.565	-0.197	0.84416
photo	29.793	11.546	2.580	0.00991 **
question	-100.506	31.669	-3.174	0.00152 **
swf	-56.261	24.329	-2.312	0.02081 *
video	-37.784	13.955	-2.708	0.00681 **

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 206.4 on 3268 degrees of freedom  
Multiple R-squared: 0.04336, Adjusted R-squared: 0.0416  
F-statistic: 24.69 on 6 and 3268 DF, p-value: < 2.2e-16

## Shares

Call:  
lm(formula = shares ~ link + music + photo + question + swf + video)

Residuals:

Min	1Q	Median	3Q	Max
-255.0	-178.6	-32.3	0.0	6170.9

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.932	19.569	0.405	0.685
link	24.333	23.150	1.051	0.293
music	220.401	238.064	0.926	0.355
photo	231.131	22.989	10.054	<2e-16 ***
question	-7.932	63.056	-0.126	0.900
swf	-7.932	48.442	-0.164	0.870
video	247.114	27.786	8.894	<2e-16 ***

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 410.9 on 3268 degrees of freedom  
Multiple R-squared: 0.06754, Adjusted R-squared: 0.06583  
F-statistic: 39.45 on 6 and 3268 DF, p-value: < 2.2e-16

Photo

Video

# Regression\_Model ( Independent Variables v.s Shares, Likes, Comments )

## Emotion

```
Call:
lm(formula = C2B_likes ~ anticipation + disgust + negative +
    trust)

Residuals:
    Min      1Q Median      3Q     Max 
-2162.5 -806.3 -419.3  430.8 6507.8 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  641.3     914.9   0.701  0.48759  
anticipation 1862.3    678.9   2.743  0.00924 ** 
disgust       6770.2   2163.7   3.129  0.00336 ** 
negative     -2640.1   1000.9  -2.638  0.01203 *  
trust        -1371.2   682.6  -2.009  0.05169 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1618 on 38 degrees of freedom  
Multiple R-squared: 0.2802, Adjusted R-squared: 0.2044  
F-statistic: 3.698 on 4 and 38 DF, p-value: 0.01225

Anticipation

Trust

Adventure

Luxury

Disgust

Negative

Affordable

## Topic

```
Call:
lm(formula = shares ~ adventure + luxury + affordable)

Residuals:
    Min      1Q Median      3Q     Max 
-314.63 -92.29 -42.21  62.22 826.67 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  51.82     52.82  -0.981  0.33258  
adventure   240.06    87.69   2.738  0.00928 ** 
luxury      148.59    86.57   1.716  0.09402 .  
affordable  168.84    82.76   2.040  0.04815 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 224 on 39 degrees of freedom  
Multiple R-squared: 0.3353, Adjusted R-squared: 0.2841  
F-statistic: 6.557 on 3 and 39 DF, p-value: 0.001071

## Time

```
Call:
lm(formula = C2B_likes ~ afternoon + morning + night)

Residuals:
    Min      1Q Median      3Q     Max 
-2698   -983   -584   -179 101540 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 653.06    185.05   3.529 0.000423 *** 
afternoon   688.98   202.77   3.398 0.000687 *** 
morning    2045.20   293.49   6.969 3.86e-12 *** 
night      65.35    204.59   0.319 0.749422 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 3173 on 3271 degrees of freedom  
Multiple R-squared: 0.02465, Adjusted R-squared: 0.02376  
F-statistic: 27.56 on 3 and 3271 DF, p-value: < 2.2e-16

Morning

Afternoon

## Type

```
Call:
lm(formula = shares ~ link + music + photo + question + swf +
    video)

Residuals:
    Min      1Q Median      3Q     Max 
-255.0 -178.6 -32.3   0.0 6170.9 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.932    19.569   0.405  0.685  
link        24.333   23.150   1.051  0.293  
music       220.401  238.064   0.926  0.355  
photo       231.131  22.989  10.054 <2e-16 *** 
question    -7.932   48.442   -0.164  0.870  
swf         -7.932   27.786   8.894 <2e-16 *** 
video       247.114  27.786   8.894 <2e-16 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 410.9 on 3268 degrees of freedom  
Multiple R-squared: 0.06754, Adjusted R-squared: 0.06583  
F-statistic: 39.45 on 6 and 3268 DF, p-value: < 2.2e-16

Photo

Video

# Audi\_KOL Top10

Number	Company_urlname	Post_id	Popularity
214	audi	10150580131986400	40399
221	audi	10150562899821400	27321
232	audi	10150520628186400	25532
239	audi	10150501370041400	24854
229	audi	10150535554181400	23234
220	audi	10150564853201400	21010
222	audi	10150560921141400	20943
230	audi	10150533644256400	17874
225	audi	235558243204199	17566
224	audi	10150554472006400	16581

---

# *Recommendation For Audi*

# How to boost social media sentiment in 2012 ?

---

## Insights

- **Emotion :** Anticipation/Disgust
- **Topic :** Adventure/Love/Memorable
- **Time :** Morning
- **Post Type:** Video
- **Year:** 2013/2011
- **Functionalities:**  
allroads/race/ski/wifi/electric/  
foursquare/leds//quattro

## Positive group

1. Create Events on key Topic (UGC)
2. Share “Audi Moments” of Top10 KOL
3. Post in the morning
4. Make Video Ad a priority
5. Review 2011&2012 version
6. Highlight “hot” functionalities

## Negative group

1. Stop the “Vampire” campaign
2. Share safety test result
3. Enhance customers service support

---



**Thank you!**

**Q&A**