



ΔΙΕΘΝΕΣ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΤΗΣ ΕΛΛΑΔΟΣ

## ΠΡΟΒΛΕΨΕΙΣ ΕΣΟΔΩΝ ΠΕΛΑΤΩΝ ΤΟΥ GOOGLE ANALYTICS

Google Analytics Customer Revenue Prediction

ΠΟΖΙΔΗΣ ΑΡΙΣΤΟΤΕΛΗΣ

ΑΕΜ : 3997

ΤΣΙΜΠΙΡΗΣ ΑΛΚΙΒΙΑΔΗΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ



Who am I?



The idea  
behind?

The target?



# Επιστήμη Των Δεδομένων

- Δεδομένα
- Αποφάσεις
- Εξέλιξη



# Εξόρυξη Δεδομένων

- Ανακάλυψη γνώσης
- Αυτόματη ανάλυση μεγάλων ποσοτήτων
- Χρήση αλγορίθμων και στατιστικής
- Κατανοητή δομή προς τον άνθρωπο
- Τεχνητή νοημοσύνη και μηχανική μάθηση



# Δεδομένα

- ❖ Ανακάλυψη γνώσης
- ❖ Σύνολα δεδομένων
- ❖ Επεξεργασία
- ❖ Καθαρισμός



# Αρχή Pareto

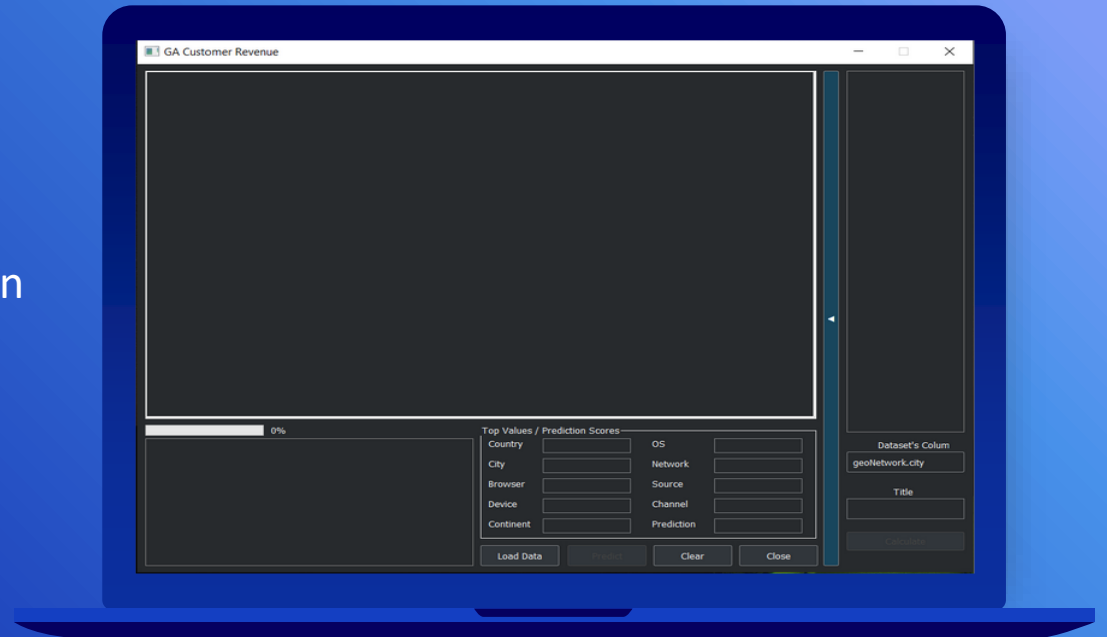
- Vilfredo Pareto
- Πανεπιστήμιο της Λοζάνης 1896
- Αρχή ακεραιότητας παραγόντων
- Κατανομή
- Αποδοτικότητα



# Εφαρμογή

## Τεχνολογίες

- Python
- Pandas
- NumPy
- Matplotlib
- LightGBM
- Scikit-Learn
- PyQt



Graph  
Tabs

Graph

Loading Bar  
& Progress  
Console

The screenshot shows a web application interface with several components:

- Graph Tabs:** A row of tabs at the top including 'Revenue Graph', 'Device Browser', 'Revenue X Browser', 'Channel Grouping', 'OS', 'Transaction by OS', 'Cities', 'Countries', and 'SubContinen'.
- Graph:** A scatter plot titled 'The 80/20 rule' with 'TransactionRevenue' on the y-axis (0 to 25) and 'index' on the x-axis (0 to 1750000). The plot shows a dense vertical line of points at the right end of the x-axis and a horizontal line at the bottom.
- Loading Bar & Progress Console:** A green progress bar at 100% and a text area containing: 'Loading Train Data (~20GB). Started at : 17:53:35.239317. Loading cleaned train dataset... Finished at : 17:59:25.366143. Train data shape : (1708337, 49). Successfully loaded train data.'
- Top Values & Prediction Scores:** A table with columns for Country, City, Browser, Device, Continent, OS, Network, Source, Channel, and Prediction. Values include 'United states', 'Mountain view', 'Chrome', 'Desktop', 'Northern america', 'Windows', 'Comcast.net', 'Google', and 'Organic search'.
- Control Buttons:** 'Load Data', 'Predict', 'Clear', and 'Close' buttons at the bottom of the prediction scores table.
- Results Console:** A text area on the right showing 'Directory Plots Already Exists' and 'Percentage of Browser usage:' followed by a list of browsers and their counts.
- Pie Chart Creation:** A section on the right showing 'Percentage of Channel Grouping used:' followed by a list of channel groups and their counts.

Results  
Console

Pie Chart  
Creation

Top Values &  
Prediction  
Scores

Control  
Buttons



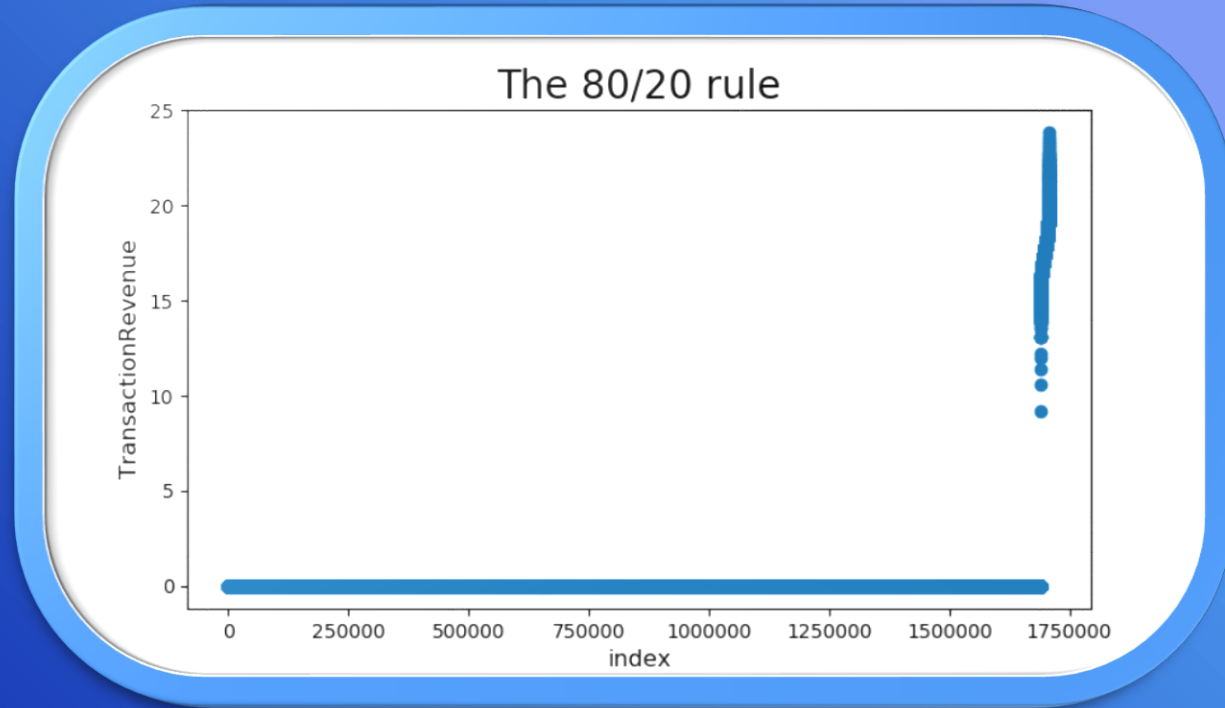
# Αποτελέσματα

Σύνολο χρηστών :

1.750.000

Χρήστες που επέφεραν  
κέρδος :

402.500

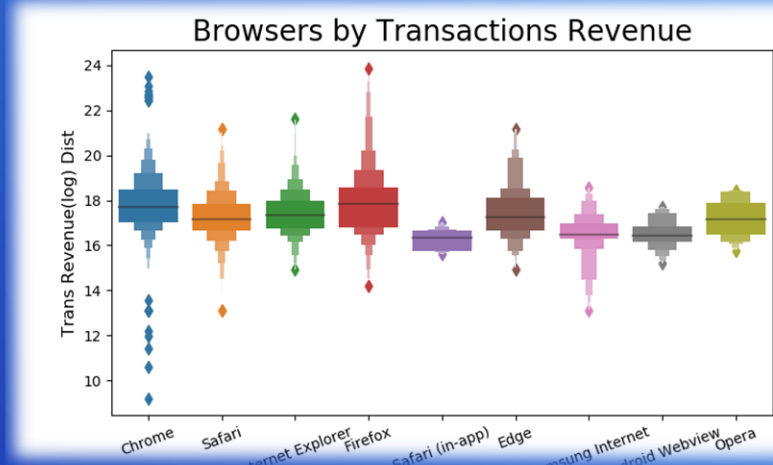
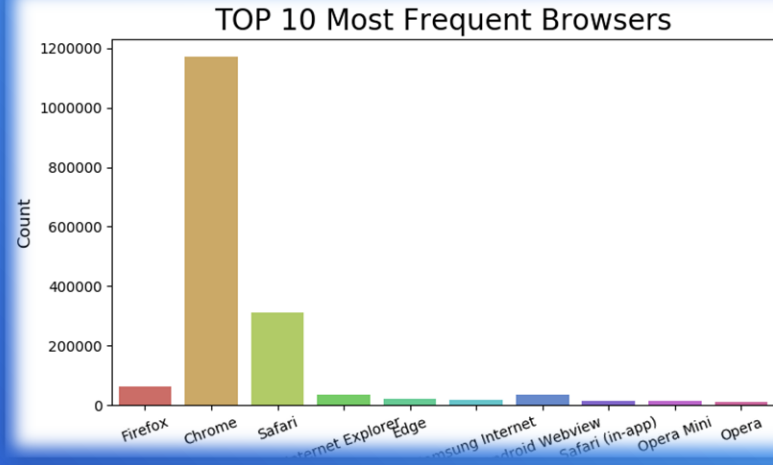


## Κορυφαίοι φυλλομετρητές :

1. Chrome
2. Safari
3. Firefox

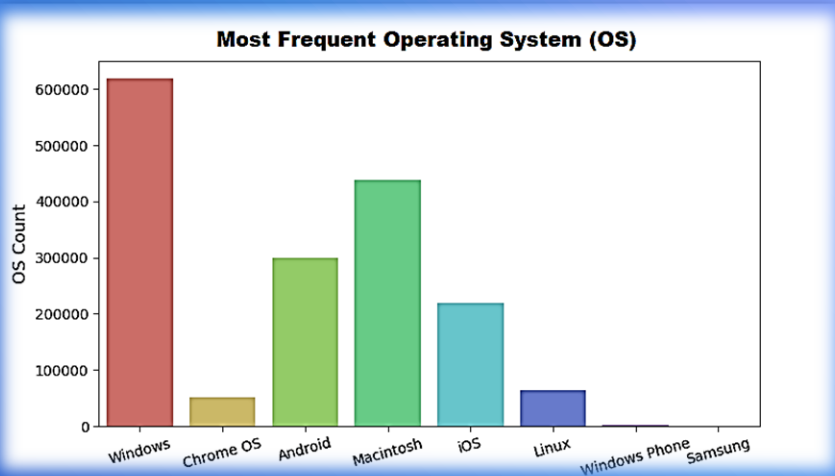
## Συναλλαγές ανά φυλλομετρητή :

1. Firefox
2. Chrome
3. Internet Explorer



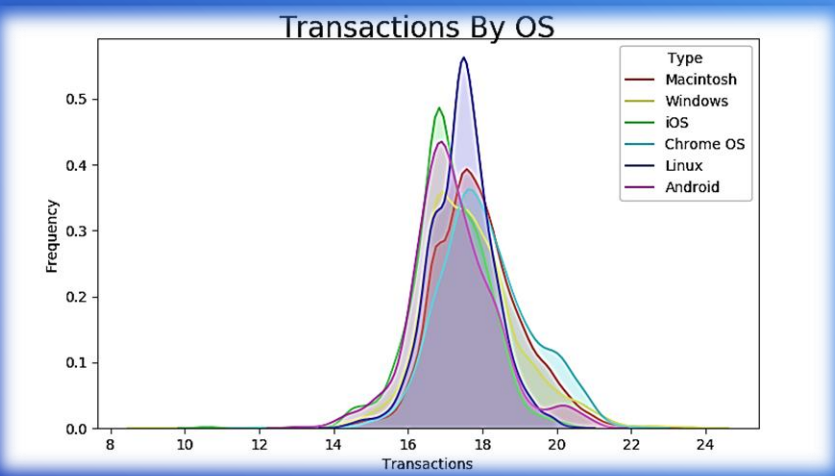
# Κορυφαία Λειτουργικά Συστήματα :

1. Windows
2. Macintosh
3. Android



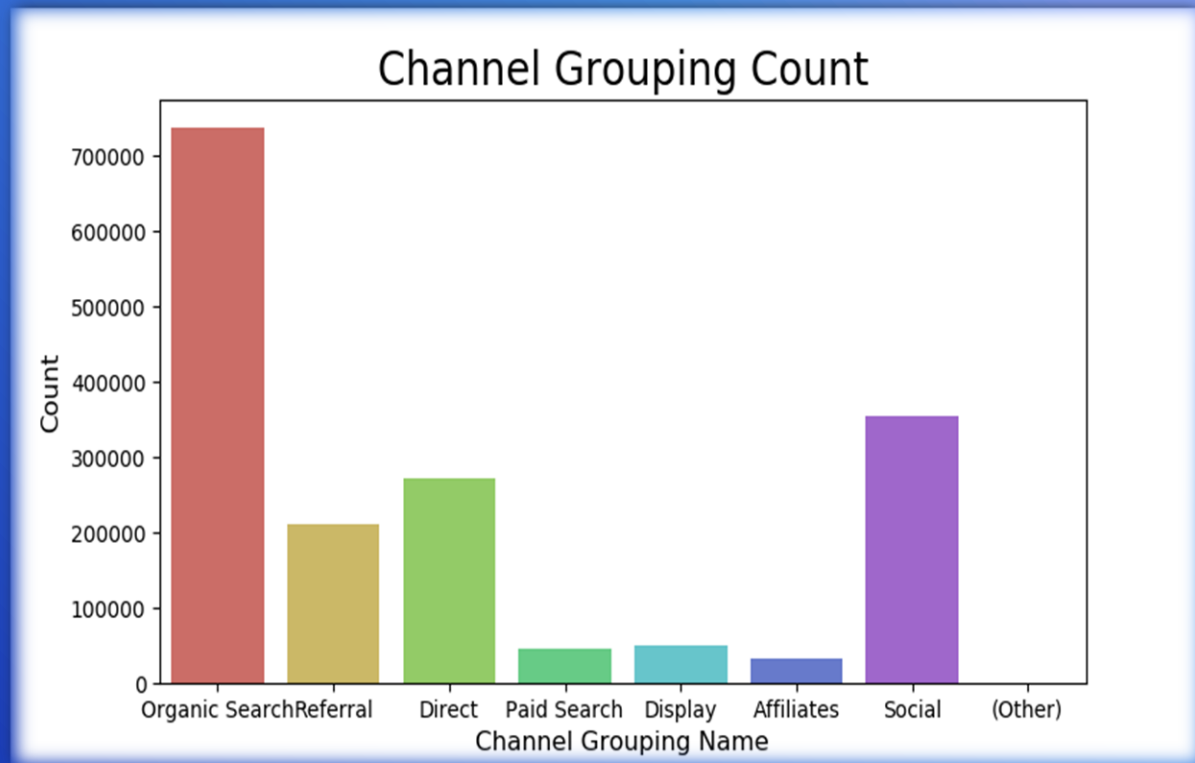
# Συναλλαγές ανά Λειτουργικό Σύστημα :

1. Linux
2. iOS
3. Android



## Κορυφαία Κανάλια :

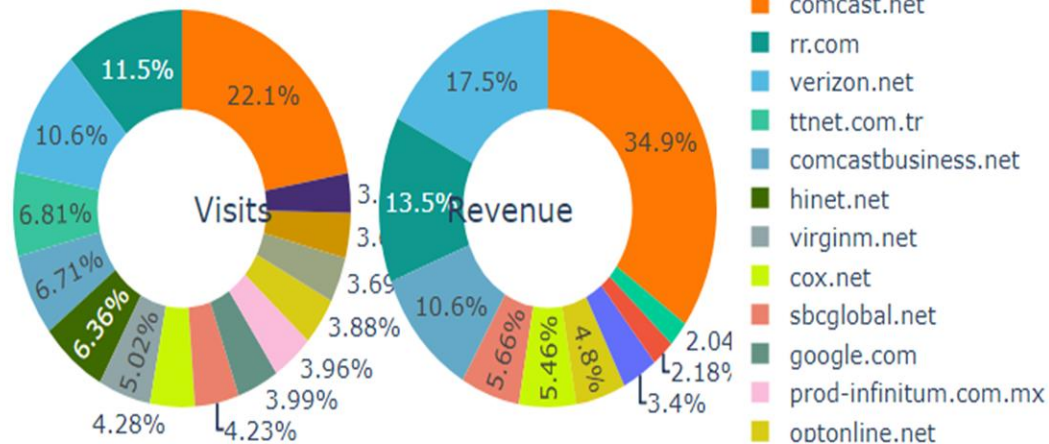
1. Organic Search
2. Social
3. Direct
4. Referral



## Κορυφαίοι Τομείς Δικτύου :

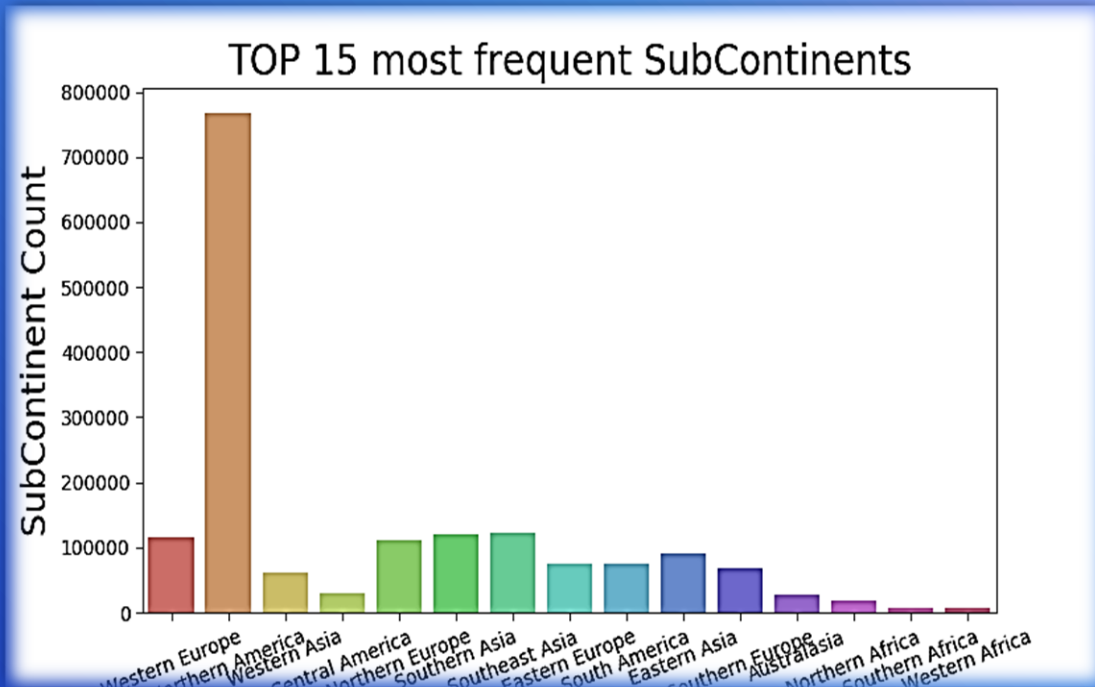
1. **comcast.net**
2. verizon.net
3. rr.com
4. comcastbusiness.net

Network Domain



Κορυφαίοι Υποήπειροι :

1. Northern America
2. Western Europe
3. Southern Asia
4. Southeast Asia

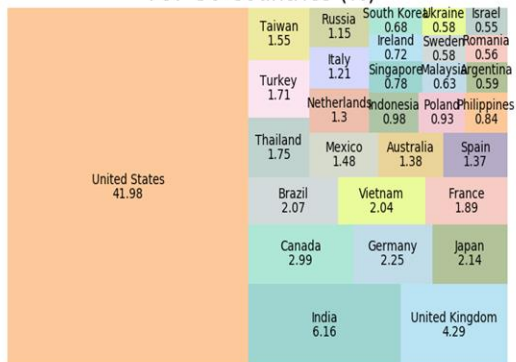


# Κορυφαίες Χώρες :

1. United States
2. India
3. United Kingdom
4. Canada
5. Germany
6. Japan



'TOP 30 Countries (%)



# Κορυφαίες Πόλεις :

1. Mountain View
2. New York
3. San Francisco
4. Sunnyvale
5. London
6. San Jose



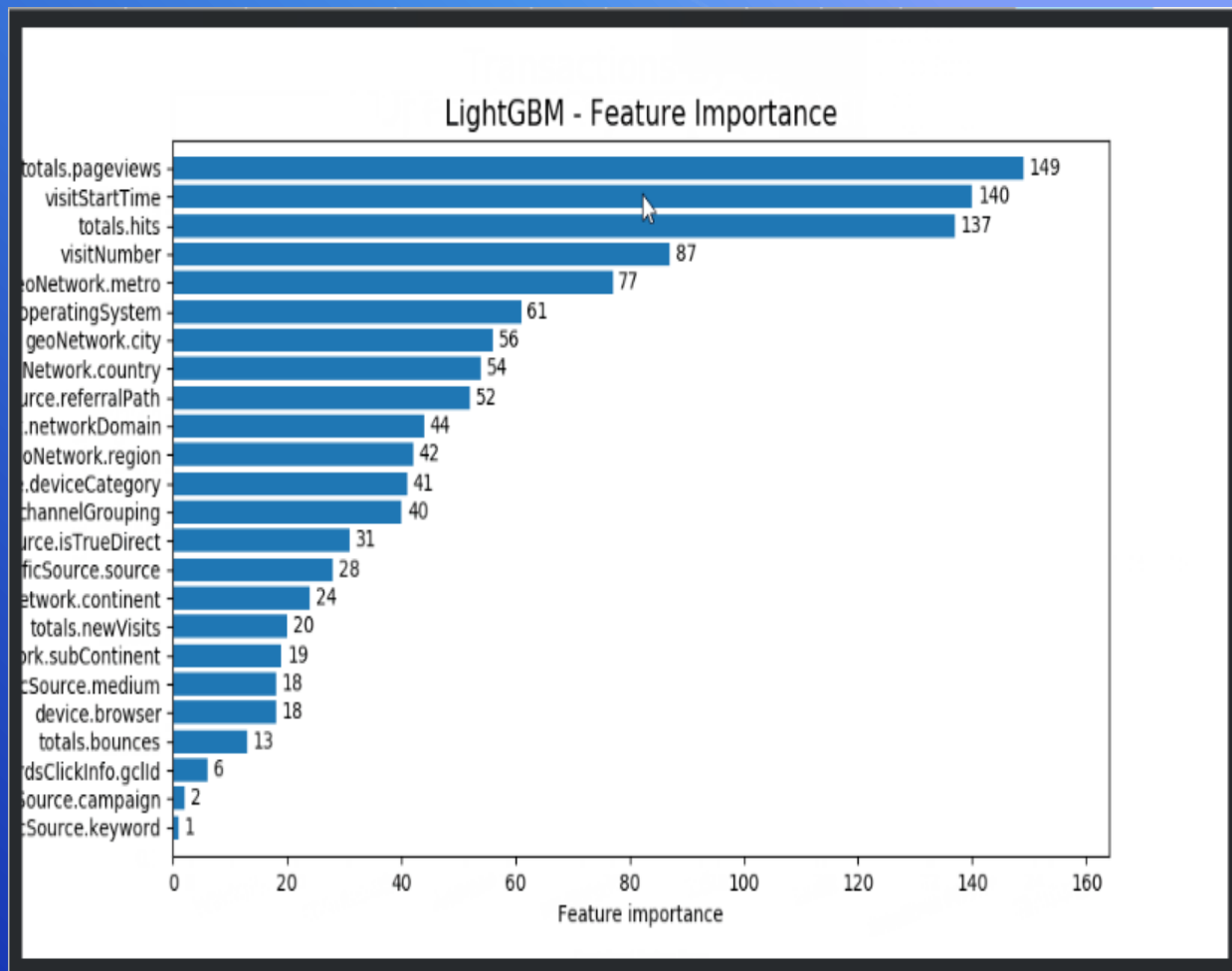
'TOP 30 Cities (%)

San Francisco 2.16	Seattle 0.56	Chennai 0.47	Santa Clara 0.38	New Delhi 0.36	Houston 0.33	Palo Alto 0.34
	Istanbul 0.58	Paris 0.47	Singapore 0.38	Jakarta 0.38	Hong Kong 0.37	
New York 2.9	Toronto 0.6	Hyderabad 0.49	Sao Paulo 0.46	Austin 0.43	Hanoi 0.41	
	Ho Chi Minh City 0.61	Dublin 0.55	Sydney 0.52	Mumbai 0.5		
Mountain View 4.34	Los Angeles 1.0	Chicago 0.89	Bangkok 0.73	Bengaluru 0.67		
	Sunnyvale 1.63	London 1.38	San Jose 1.18			



## Σημαντικότητα Στηλών:

1. total.pageviews
2. visitStartTime
3. totals.hits
4. geoNetwork.metro
5. operatingSystem





# Επέκταση εφαρμογής και συμπέρασμα.

Σας  
ευχαριστώ!

Αριστοτέλης Ποζίδης  
arist.poz@gmail.com

