

Διαχείριση Μεγάλων Δεδομένων 1^η Προγραμματιστική Εργασία

Διδάσκουσα:
Π. Ραυτοπούλου

Παράδοση μέχρι: **Παρασκευή 22/12/2023 ώρα 23.59**
Προσωπική εξέταση: εξεταστική Φεβρουαρίου
(η ακριβής ημερομηνία θα ανακοινωθεί έγκαιρα)

ΣΗΜΑΝΤΙΚΕΣ ΣΗΜΕΙΩΣΕΙΣ:

1. Σε όλα τα αρχεία που θα παραδώσετε θα πρέπει **ΟΠΩΣΔΗΠΟΤΕ** να βάλετε τα ονόματα, τους A.M., και τα username/email των μελών της ομάδας (ομάδες 2 ή 3 ατόμων).
2. Αφού έχετε ολοκληρώσει την εργασία που θέλετε να παραδώσετε, την υποβάλετε στο eclass στο υποσύστημα «Εργασίες» (Assignments). Προσοχή: μόνο 1 άτομο από την ομάδα χρειάζεται να παραδώσει την εργασία μέσω του e-class! Η υποβολή πρέπει να γίνει **ΠΡΙΝ** την καταληκτική ημερομηνία παράδοσης. Παραδίδετε όλα τα αρχεία μαζί σε ένα συμπίεσμένο αρχείο (το οποίο θα φέρει τα ονόματα της ομάδας π.χ., RaftoroulouParadopoulos.zip).
3. Περιπτώσεις αντιγραφής θα μηδενίζονται κι οι εμπλεκόμενοι δε θα έχουν δικαίωμα παράδοσης άλλων εργασιών.
4. Η ημερομηνία παράδοσης είναι αυστηρή, και η παράδοση γίνεται μόνο μέσω του eclass και όχι με email στη διδάσκουσα. Ασκήσεις που παραδίδονται μετά τη λήξη της προθεσμίας δε γίνονται δεκτές.

Πώς να γίνω εκατομμυριούχος!

Το αμερικανικό επιχειρηματικό περιοδικό Forbes δημοσιεύεται οκτώ φορές το χρόνο και παρουσιάζει άρθρα για θέματα οικονομίας, βιομηχανίας, επενδύσεων και μάρκετινγκ. Περιλαμβάνει επίσης θέματα σχετικά με την τεχνολογία, τις επικοινωνίες, την επιστήμη, την πολιτική και το δίκαιο. Το Forbes έχει έκδοση, εκτός από τις ΗΠΑ, στην Ασία καθώς και εκδόσεις σε άλλες 27 χώρες και περιοχές σε όλο τον κόσμο. Γιατί να μας ενδιαφέρει όμως το συγκεκριμένο περιοδικό;

Το Forbes, μέσω μιας ομάδας με περισσότερους από 50 ρεπόρτερ από διάφορες χώρες, παρακολουθεί τη δραστηριότητα των πλουσιότερων ανθρώπων του κόσμου ή/και ομάδων/οικογενειών που μοιράζονται πλούτο, και, από το Μάρτιο του 1987, δίνει κάθε χρόνο στη δημοσιότητα μια ετήσια κατάταξη ανθρώπων που είναι δισεκατομμυριούχοι, δηλαδή που έχουν καθαρή περιουσία ίση ή μεγαλύτερη από 1 δισεκατομμύριο δολάρια. Η λίστα αυτή ονομάζεται The World's Billionaires! Η συνολική καθαρή αξία κάθε ατόμου στη λίστα εκτιμάται και αναφέρεται σε δολάρια Ηνωμένων Πολιτειών, με βάση τα τεκμηριωμένα περιουσιακά τους στοιχεία, τη λογιστική για το χρέος, και άλλους παράγοντες. Οι βασιλιάδες και οι δικτάτορες, των οποίων ο πλούτος προέρχεται από τις θέσεις τους, εξαιρούνται από αυτές τις λίστες. Εξαιρούνται επίσης, από αυτή η κατάταξη όσοι έχουν περιουσία που δεν είναι δυνατό να εξακριβωθεί πλήρως.

Εγκυκλοπαιδικά μιλώντας, το 2018, ο ιδρυτής της Amazon, Τζεφ Μπέζος, κατατάχθηκε στην κορυφή για πρώτη φορά και έγινε ο πρώτος άνθρωπος με πάνω από 100 δισεκατομμύρια δολάρια που συμπεριλήφθηκε στην λίστα, ξεπερνώντας τον ιδρυτή της Microsoft, Μπιλ Γκέιτς, ο οποίος βρέθηκε στην κορυφή της λίστας για τα 18 από τα 24 προηγούμενα έτη. Το 2022, μετά από τέσσερα συνεχόμενα χρόνια στην κορυφή της λίστας, ο Ίλον Μασκ (επικεφαλής των Twitter, SpaceX, Tesla Inc., κ.α.) ξεπέρασε τον Μπέζος. Το 2023, ο Γάλλος επιχειρηματίας Bernard Arnault της Luis Vuitton έγινε ο πρώτος Γάλλος που έφτασε στην κορυφή της λίστας.

Τι πρέπει να κάνουμε εμείς για να πλασαριστούμε στη λίστα του Forbes; Μπορούμε να ανακαλύψουμε τα μυστικά μελετώντας τα δεδομένα που ήδη έχουμε;

Σκοπός της εργασίας

Σκοπός αυτής της εργασίας είναι να μελετήσετε και να αναλύσετε ένα σύνολο δεδομένων με στατιστικά για επιτυχημένους επιχειρηματίες από όλον τον κόσμο. Τα δεδομένα που θα χρησιμοποιήσετε είναι δωρεάν και δημοσίως διαθέσιμα.

Το σύνολο δεδομένων που έχετε στα χέρια σας περιέχει πληθώρα πληροφοριών για τους εκατομμυριούχους. Ωστόσο, με τόσο τεράστιο όγκο δεδομένων, δεν είναι προφανές πώς θα μπορούσαμε κι εμείς να γίνουμε ένας από αυτούς! Στόχος σας είναι να εξάγετε ουσιαστικές γνώσεις και να απαντήσετε σε βασικά ερωτήματα σχετικά με ... τις συνήθειες των εκατομμυριούχων 😊

Το πρόβλημα που στοχεύουμε να αντιμετωπίσουμε σε αυτήν την ανάλυση είναι να κατανοήσετε τους παράγοντες που συμβάλλουν στον πλούτο και τις επιπτώσεις της ανισότητας του πλούτου. Εξετάζοντας το σύνολο δεδομένων, θα διερευνήσετε την κατανομή του πλούτου σε διαφορετικούς κλάδους και χώρες, θα αξιολογήσετε τον αντίκτυπο των κοινωνικοοικονομικών παραγόντων στον πλούτο των εκατομμυριούχων, και θα διερευνήσετε τις ανισότητες.

Τελικά, αυτή η ανάλυση μπορεί να μας δώσει πολύτιμες πληροφορίες για τη δυναμική του πλούτου, ρίχνοντας φως στις οικονομικές ανισότητες και στους πιθανούς τομείς για περαιτέρω έρευνα και πολιτική παρέμβαση.

Η εργασία θα εκπονηθεί από ομάδες των **2 ή 3 ατόμων**, θα υλοποιηθεί σε **Java**, και μπορεί να γίνει σε οποιοδήποτε λειτουργικό σύστημα (συστήνεται το GNU/Linux).

Για τις ανάγκες της εργασίας, θα κατεβάσετε και θα εγκαταστήσετε το Hadoop στο μηχανήμα σας, κατά προτίμηση σε **Single-Node Local (Standalone) Mode**. Το mode αυτό προορίζεται για debugging και έχει ευκολότερη διαδικασία εγκατάστασης, οπότε το συστήνω μιας και θα κάνετε την υλοποίηση στο μηχανήμα σας κι όχι σε κάποιο cluster. Προσοχή, θα πρέπει να κατεβάσετε και να εγκαταστήσετε την έκδοση του Hadoop που ταιριάζει με το setup του μηχανήματός σας!

Εκτός από τις διαφάνειες του μαθήματος, στις οποίες συζητήσαμε για το Hadoop και για το MapReduce, επιπλέον πληροφορίες για την ακριβή λειτουργία τους μπορείτε να βρείτε στα Έγγραφα και στους Συνδέσμους (όλα διαθέσιμα στο eclass του μαθήματος).

Λειτουργικότητα του συστήματος

Τα δεδομένα είναι αποθηκευμένα σε ένα αρχείο κειμένου csv (comma-separated values) και αποτελούν ένα (μικρό) δείγμα για να μπορείτε να τα διαχειριστείτε από τους προσωπικούς σας υπολογιστές. Τα δεδομένα αφορούν σε 2641 πολύ πλούσιους ανθρώπους (σύμφωνα με την κατάταξη του Forbes για το 2023). Τα δεδομένα είναι διαθέσιμα από τα Έγγραφα στο eclass του μαθήματος.

Κάθε γραμμή/εγγραφή σε αυτό το αρχείο αντιστοιχεί σε έναν εκατομμυριούχο και περιλαμβάνει τα εξής 35 χαρακτηριστικά (attributes):

1. **rank**: The ranking of the billionaire in terms of wealth
2. **finalWorth**: The final net worth (in millions) of the billionaire in U.S. dollars
3. **category**: The category or industry in which the billionaire's business operates
4. **personName**: The full name of the billionaire
5. **age**: The age of the billionaire
6. **country**: The country in which the billionaire resides
7. **city**: The city in which the billionaire resides
8. **source**: The source of the billionaire's wealth
9. **industries**: The industries associated with the billionaire's business interests
10. **countryOfCitizenship**: The country of citizenship of the billionaire
11. **organization**: The name of the organization or company associated with the billionaire
12. **selfMade**: Indicates whether the billionaire is self-made (True/False)
13. **status**: "D" represents self-made billionaires (Founders/Entrepreneurs) and "U" indicates inherited or unearned wealth
14. **gender**: The gender of the billionaire
15. **birthDate**: The birthdate of the billionaire
16. **lastName**: The last name of the billionaire
17. **firstName**: The first name of the billionaire
18. **title**: The title or honorific of the billionaire
19. **date**: The date of data collection
20. **state**: The state in which the billionaire resides
21. **residenceStateRegion**: The region or state of residence of the billionaire
22. **birthYear**: The birth year of the billionaire

23. **birthMonth**: The birth month of the billionaire
24. **birthDay**: The birth day of the billionaire
25. **cpi_country**: Consumer Price Index (CPI) for the billionaire's country
26. **cpi_change_country**: CPI change for the billionaire's country
27. **gdp_country**: Gross Domestic Product (GDP) for the billionaire's country
28. **gross_tertiary_education_enrollment**: Enrollment in tertiary education in the billionaire's country
29. **gross_primary_education_enrollment_country**: Enrollment in primary education in the billionaire's country
30. **life_expectancy_country**: Life expectancy in the billionaire's country
31. **tax_revenue_country_country**: Tax revenue in the billionaire's country
32. **total_tax_rate_country**: Total tax rate in the billionaire's country
33. **population_country**: Population of the billionaire's country
34. **latitude_country**: Latitude coordinate of the billionaire's country
35. **longitude_country**: Longitude coordinate of the billionaire's country

Παρακάτω βρείτε ένα παράδειγμα τριών εγγραφών:

```
1,211000,Fashion & Retail,Bernard Arnault & family,74,France,Paris,LVMH,Fashion & Retail,France,LVMH Moët Hennessy Louis  
Vuitton,FALSE,U,M,3/5/1949 0:00,Arnault,Bernard,Chairman and CEO,4/4/2023  
5:01,,,1949,3,5,110.05,1.1,"$2,715,518,274,227",65.6,102.5,82.5,24.2,60.7,67059887,46.227638,2.213749  
2,180000,Automotive,Elon Musk,51,United States,Austin,"Tesla, SpaceX",Automotive,United States,Tesla,TRUE,D,M,6/28/1971  
0:00,Musk,Elon,CEO,4/4/2023  
5:01,Texas,South,1971,6,28,117.24,7.5,"$21,427,700,000,000",88.2,101.8,78.5,9.6,36.6,328239523,37.09024,-95.712891  
3,114000,Technology,Jeff Bezos,59,United States,Medina,Amazon,Technology,United States,Amazon,TRUE,D,M,1/12/1964  
0:00,Bezos,Jeff,Chairman and Founder,4/4/2023  
5:01,Washington,West,1964,1,12,117.24,7.5,"$21,427,700,000,000",88.2,101.8,78.5,9.6,36.6,328239523,37.09024,-95.712891
```

Προσέξτε ότι τα δεδομένα που αφορούν σε μια εγγραφή (δηλαδή σε έναν εκατομμυριούχο) χωρίζονται μεταξύ τους με κόμμα. Παρατηρείστε επίσης, ότι μπορεί κάποια εγγραφή να έχει κενά κάποιο ή κάποια χαρακτηριστικά. Για παράδειγμα, η πρώτη εγγραφή στο παραπάνω παράδειγμα δεν έχει τιμές για τα χαρακτηριστικά 20 (state) και 21 (residenceStateRegion). Δείτε με προσοχή στο αρχείο των δεδομένων ενδεικτικές τιμές των χαρακτηριστικών για να καταλάβετε σε τι ακριβώς αφορούν τα συγκεκριμένα δεδομένα. Αποφασίστε πώς θέλετε να χειριστείτε τις κενές τιμές σε κάθε χαρακτηριστικό.

Οποιαδήποτε παρατήρηση, απόφασή σας ή/και (προ-)επεξεργασία των δεδομένων θα πρέπει να περιγράφεται με σαφήνεια στην γραπτή αναφορά σας. Η (προ-)επεξεργασία, αν κάνετε, των δεδομένων θα πρέπει να γίνει με κώδικα ή/και scripts (ανάλογα με την υλοποίηση μπορεί να λάβει bonus ως και 5%).

Χρησιμοποιείτε τα δεδομένα για να κάνετε τους παρακάτω υπολογισμούς και καταγράψτε σε αντίστοιχα αρχεία εξόδου τα αποτελέσματα που θα πάρετε. Σημειώστε ότι οι εγγραφές που εμφανίζονται στα παραδείγματα που ακολουθούν είναι τυχαίες και δεν αντιστοιχούν σε πραγματικούς υπολογισμούς.

Μελέτη των χαρακτηριστικών [20%]

Γράψτε στην αναφορά σας πόσα και ποια από τα χαρακτηριστικά είναι numeric (πείτε αν είναι interval ή ratio-scaled) και πόσα και ποια είναι nominal (επιστημάνετε τα binary και τα ordinal). Επιλέξτε ένα numeric κι ένα nominal (όχι όμως binary) χαρακτηριστικό και, για κάθε ένα από αυτά, κάντε τα παρακάτω.

(α) Γράψτε στην αναφορά σας τον **ψευδοκώδικα** για τη λειτουργία Map και για τη λειτουργία Reduce (και τον Combiner, αν χρειάζεται) για να βρείτε το εύρος (range) των τιμών του. Δώστε ένα **παράδειγμα** και μια **σηματική εκτέλεση** που να εξηγήει τη λογική πίσω από τον ψευδοκώδικά σας.

(β) Υλοποιήστε στο Hadoop τον σχεδιασμό που κάνατε παραπάνω. Το αρχείο εξόδου θα πρέπει να περιέχει το εύρος των τιμών του χαρακτηριστικού.

Επιστημάνετε στην αναφορά σας, αν υπάρχει, τη διαφορά στην αντιμετώπιση των παραπάνω χαρακτηριστικών. Φτιάξτε τουλάχιστον μια γραφική απεικόνιση από τα αποτελέσματα που πήρατε.

Σημείωση: Για το nominal χαρακτηριστικό, προτείνετε στην αναφορά σας έναν τρόπο που θα μπορούσε να αποτυπώνει το εύρος των τιμών στο σύνολο των δεδομένων και υλοποιήστε την ιδέα σας.

Σε ποια ηλικία θα καταφέρω να γίνω πλούσιος; [20%]

(α) Θεωρήστε διαστήματα με βήμα δέκα (10) από τα 15 ως τα 104 έτη και γράψτε στην αναφορά σας τον **ψευδοκώδικα** για τη λειτουργία Map και για τη λειτουργία Reduce (και τον Combiner, αν χρειάζεται) για να βρείτε το πλήθος των εκατομμυριούχων ανά ηλικιακό διάστημα. Δώστε ένα **παράδειγμα** και μια **σηματική εκτέλεση** που να εξηγήει τη λογική πίσω από τον ψευδοκώδικά σας.

(β) Υλοποιήστε στο Hadoop τον σχεδιασμό που κάνατε παραπάνω. Το αρχείο εξόδου θα πρέπει να περιέχει τόσες γραμμές όσα και τα διαφορετικά ηλικιακά διαστήματα και σε κάθε γραμμή να εμφανίζονται (i) το ηλικιακό διάστημα και (ii) το πλήθος των εκατομμυριούχων που ανήκουν σε αυτό.

Δείτε παρακάτω ένα παράδειγμα:

```
15-24 5
25-34 20
...
```

Φτιάξτε ένα ιστόγραμμα από τα αποτελέσματα που πήρατε.

Σε ποια χώρα να ζω τέλος πάντων για να γίνω πλούσιος; [20%]

(α) Γράψτε στην αναφορά σας τον **ψευδοκώδικα** για τη λειτουργία Map και για τη λειτουργία Reduce (και τον Combiner, αν χρειάζεται) για να βρείτε σε ποιες χώρες συγκεντρώνεται ο περισσότερος πλούτος και αν ο πλούτος επηρεάζεται από την αγοραστική δύναμη των καταναλωτών της χώρας. Ως αποτέλεσμα θέλουμε την αγοραστική δύναμη για κάθε χώρα, το άθροισμα του πλούτου, και το πλήθος των εκατομμυριούχων. Δώστε ένα **παράδειγμα** και μια **σχηματική εκτέλεση** που να εξηγεί τη λογική πίσω από τον ψευδοκώδικά σας.

(β) Υλοποιήστε στο Hadoop τον σχεδιασμό που κάνατε παραπάνω. Το αρχείο εξόδου θα πρέπει να περιέχει (i) το όνομα της χώρας (**country**), (ii) την αγοραστική δύναμη των καταναλωτών της συγκεκριμένης χώρας (**cri_country**), (iii) το άθροισμα του πλούτου (**finalWorth**) όλων των εκατομμυριούχων που ζουν σε αυτή τη χώρα, και (iv) το πλήθος των εκατομμυριούχων.

Δείτε παρακάτω ένα παράδειγμα:

```
United States 117.24 4575100 754
China 125.08 1770400 504
Greece 101.87 9100 3
...
```

Απεικονίστε τα αποτελέσματά σας με ένα scatter plot και εξηγήστε αν και πώς ο πλούτος εξαρτάται από/συσχετίζεται με τη χώρα. Σκεφθείτε τι θα αντιπροσωπεύει κάθε ένας από τους άξονες του γραφήματος για να έχει νόημα η απεικόνιση.

Με ποιο τομέα πρέπει να ασχοληθώ για να πιάσω την καλή; [30%]

(α) Γράψτε στην αναφορά σας τον **ψευδοκώδικα** για τη λειτουργία Map και για τη λειτουργία Reduce (και τον Combiner, αν χρειάζεται) για να βρείτε το ποσοστό που κατέχει κάθε κατηγορία, στην οποία δραστηριοποιούνται οι εκατομμυριούχοι, επί του συνολικού πλούτου. Για κάθε κατηγορία θέλουμε να γνωρίζουμε το ποσοστό επί του συνολικού πλούτου, το πλήθος των εκατομμυριούχων, το άθροισμα του πλούτου τους, τον μέσο όρο (mean) του πλούτου τους, και τον ισχυρότερο εκπρόσωπο της κατηγορίας. Δώστε ένα **παράδειγμα** και μια **σχηματική εκτέλεση** που να εξηγεί τη λογική πίσω από τον ψευδοκώδικά σας.

(β) Υλοποιήστε στο Hadoop τον σχεδιασμό που κάνατε παραπάνω. Το αρχείο εξόδου θα πρέπει να περιέχει τις κατηγορίες και σε κάθε γραμμή να εμφανίζονται (i) η κατηγορία (**category**), (ii) το ποσοστό που κατέχει επί του συνολικού πλούτου, (iii) το πλήθος των εκατομμυριούχων που δραστηριοποιείται σε αυτή, (iv) το άθροισμα του πλούτου τους (**finalWorth**), (v) τον μέσο όρο του πλούτου τους, και (vi) τον ισχυρότερο εκπρόσωπό της (σε όρους πλούτου).

Δείτε παρακάτω ένα παράδειγμα:

```
1 Technology 15% 307 1862000 6065,15 Jeff Bezos
2 Fashion & Retail 14% 259 1659900 6408,88 Bernard Arnault & family
2 Finance & Investments 13% 363 1586400 4370,25 Warren Buffett
3
...
```

Φτιάξτε δυο γραφικές απεικονίσεις από τα αποτελέσματα που πήρατε.

Αν στο αρχείο αποτελεσμάτων έχετε και τη σειρά κατάταξης, όπου οι κατηγορίες πρέπει να είναι ταξινομημένες σε φθίνουσα σειρά ως προς το ποσοστό επί του συνολικού πλούτου, τότε λαμβάνετε bonus 5%. Σε περίπτωση που υπάρχουν παραπάνω από μια κατηγορίες με το ίδιο ποσοστό, τότε πιο ψηλά στην κατάταξη θεωρείστε την κατηγορία με τον μεγαλύτερο πλήθος εγγραφών.

Σημείωση: ποσοστό = $100 * \text{άθροισμα πλούτου κατηγορίας} / \text{άθροισμα όλου του πλούτου}$

Αναφορά [10%]

Η αναφορά σας θα πρέπει να έχει έκταση τουλάχιστον πέντε σελίδες (χωρίς το εξώφυλλο και τις αναφορές/παραπομπές) και θα πρέπει να περιέχει:

- οδηγίες για την εκτέλεση του προγράμματος (user manual),
- λεπτομέρειες της υλοποίησης που αξίζει να σημειωθούν,
- σαφή αναφορά σε κώδικα ή/και scripts τρίτων που ενδεχομένως χρησιμοποιήσατε,
- για κάθε ένα από τα ερωτήματα
 - τον ψευδοκώδικα, μαζί με ένα παράδειγμα και μια σχηματική εκτέλεση,
 - τα (20 πρώτα) αποτελέσματα,
 - τουλάχιστον μια γραφική απεικόνιση (π.χ., boxplot, bar graph, histogram, κλ.π., όπου δεν σας ζητείται με σαφήνεια εσείς θα αποφασίσετε ποια) των αποτελεσμάτων (συστήνεται για τις απεικονίσεις η χρήση Gnuplot, Tableau, ή άλλου αντίστοιχου προγράμματος, και πάντως όχι Excel ούτε OpenOffice), και
 - τα σχόλιά σας σε σχέση με τα αποτελέσματα (π.χ., τι αναμένατε και τι πήρατε ως αποτέλεσμα, ποια είναι η κατανομή των τιμών ενός χαρακτηριστικού, πώς συσχετίζονται τα χαρακτηριστικά).

Μην συμπεριλάβετε στην αναφορά σας μέρος ή σύνολο του κώδικα, εκτός αν αφορά σε κάποια σύντομη επεξήγηση!

Bonus υλοποίηση

Σε συνεννόηση με τη διδάσκουσα μπορείτε να πάρετε μέχρι 20% bonus για επιπλέον λειτουργικότητα ή υπολογισμούς που θα υλοποιήσετε. Κάποια ενδεικτικά αναφέρονται στις προηγούμενες σελίδες. Άλλα παραδείγματα είναι, να υλοποιήσετε τα παραπάνω ερωτήματα σε Pig Latin (θα πρέπει στην αναφορά να υπάρχει σύγκριση των αποτελεσμάτων και του χρόνου εκτέλεσης σε σχέση με την Java υλοποίηση), να διερευνήσετε την κατανομή του πλούτου σε διαφορετικούς τομείς και περιοχές βρίσκοντας τη σχέση μεταξύ κατάταξης (rate) και αξίας (finalWorth) ανά κατηγορία και χώρα, να διερευνήσετε το ποσοστό των αυτοδημιούργητων ανά κατηγορία, κ.α.

Παραδοτέα και βαθμολόγηση

Πλήρης θεωρείται η εργασία η οποία **υλοποιεί σωστά τις βασικές απαιτήσεις** που περιγράφονται παραπάνω. Ασκήσεις που υλοποιούν μόνο ένα μέρος των βασικών απαιτήσεων λαμβάνουν και αντίστοιχο μέρος του βαθμού.

Τα παραδοτέα της εργασίας είναι:

- τυχόν scripts ή κώδικας τρίτων που χρησιμοποιήσατε,
- τα αρχεία πηγαίου κώδικα (συμπεριλαμβανομένων βιβλιοθηκών, αν χρειαστήκατε),
- τα εκτελέσιμα αρχεία/jars (αν προκύψουν),
- τα αρχεία με τα αποτελέσματα για κάθε ένα από τα παραπάνω ερωτήματα,
- η γραπτή αναφορά.

Καλή επιτυχία!