

Προχωρημένα Θέματα Βάσεων Δεδομένων

Εξαμηνιαία Άσκηση

Ομάδα 96

Νικόλαος Μπέλλος, el18183

Αγγελική Εμμανουέλα Συρρή, el18811

GitHub Project

Ερωτήματα:

Κάθε ένα από τα queries έχει εκτελεστεί 10 φορές, και παρακάτω παρατίθεται ο μέσος όρος του χρόνου εκτέλεσης που έχουμε λάβει.

Q1. Να βρεθεί η διαδρομή με το μεγαλύτερο φιλοδώρημα (tip) τον Μάρτιο και σημείο άφιξης το "Battery Park".

| | 1 Worker | 2 Workers |
|----------------|--------------|--------------|
| Execution Time | 13.02150 sec | 12.45035 sec |

| Feature | Value |
|-----------------------|---------------------|
| VendorID | 2 |
| tpep_pickup_datetime | 2022-03-17 12:27:47 |
| tpep_dropoff_datetime | 2022-03-17 12:27:58 |
| passenger_count | 1.0 |
| trip_distance | 0.0 |
| RatecodeID | 1.0 |
| store_and_fwd_flag | N |
| PULocationID | 12 |

| | |
|-----------------------|--------------|
| DOLocationID | 12 |
| payment_type | 1 |
| fare_amount | 2.5 |
| extra | 0.0 |
| mta_tax | 0.5 |
| tip_amount | 40.0 |
| tolls_amount | 0.0 |
| improvement_surcharge | 0.3 |
| total_amount | 45.8 |
| congestion_surcharge | 2.5 |
| airport_fee | 0.0 |
| LocationID | 12 |
| Borough | Manhattan |
| Zone | Battery Park |
| service_zone | Yellow Zone |

Q2. Να βρεθεί, για κάθε μήνα, η διαδρομή με το υψηλότερο ποσό στα διόδια.

| | | |
|----------------|--------------|-------------|
| | 1 Worker | 2 Workers |
| Execution Time | 13.19553 sec | 9.56669 sec |

| Feature | January | February | March | April | May | June |
|-----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| VendorID | 1 | 1 | 1 | 1 | 1 | 1 |
| tpcp_pickup_datetime | 2022-01-22 11:39:07 | 2022-02-18 02:33:30 | 2022-03-11 20:08:32 | 2022-04-29 04:31:21 | 2022-05-21 16:47:48 | 2022-06-12 16:51:46 |
| tpcp_dropoff_datetime | 2022-01-22 12:31:09 | 2022-02-18 02:35:28 | 2022-03-11 20:09:45 | 2022-04-29 04:32:30 | 2022-05-21 17:05:47 | 2022-06-12 17:56:48 |
| passenger_count | 1.0 | 1.0 | 1.0 | 2.0 | 1.0 | 9.0 |
| trip_distance | 33.4 | 1.3 | 0.0 | 0.0 | 2.4 | 22.0 |
| RatecodeID | 1.0 | 1.0 | 1.0 | 1.0 | 3.0 | 1.0 |

| | | | | | | |
|-----------------------|-------|--------|-------|--------|--------|--------|
| store_and_fwd_flag | Y | N | N | N | N | N |
| PULocationID | 70 | 265 | 265 | 249 | 239 | 142 |
| DOLocationID | 265 | 265 | 265 | 249 | 246 | 132 |
| payment_type | 4 | 1 | 1 | 3 | 3 | 2 |
| fare_amount | 88.0 | 3.0 | 2.5 | 3.0 | 31.5 | 67.5 |
| extra | 0.0 | 0.5 | 1.0 | 3.0 | 0.0 | 2.5 |
| mta_tax | 0.5 | 0.5 | 0.5 | 0.5 | 0.0 | 0.5 |
| tip_amount | 0.0 | 19.85 | 48.0 | 0.0 | 0.0 | 0.0 |
| tolls_amount | 193.3 | 95.0 | 235.7 | 911.87 | 813.75 | 800.09 |
| improvement_surcharge | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| total_amount | 282.1 | 119.15 | 288.0 | 918.67 | 845.55 | 870.89 |
| congestion_surcharge | 0.0 | 0.0 | 0.0 | 2.5 | 0.0 | 2.5 |
| airport_fee | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Q3. Να βρεθεί, ανά 15 ημέρες, ο μέσος όρος της απόστασης και του κόστους για όλες τις διαδρομές με σημείο αναχώρησης διαφορετικό από το σημείο άφιξης.

| | 1 Worker | 2 Workers |
|----------------------|---------------|---------------|
| Execution Time | 22.09766 sec | 18.70105 sec |
| Execution Time (RDD) | 300.30615 sec | 258.44883 sec |

| month_half | avg(trip_distance) | avg(total_amount) |
|------------|--------------------|--------------------|
| 1-a | 5.576429554927404 | 19.90405084638674 |
| 1-b | 5.097880367275346 | 19.14882164234129 |
| 2-a | 6.248888338463885 | 19.491979067238955 |
| 2-b | 5.849460516243601 | 20.18769180439039 |
| 3-a | 6.480485434052913 | 20.652278174141436 |
| 3-b | 5.556944935850653 | 21.120920554171548 |
| 4-a | 5.6793230779383554 | 21.515559094570637 |
| 4-b | 5.8003418315344 | 21.428117467105963 |

| | | |
|-----|-------------------|--------------------|
| 5-a | 6.249697852127271 | 21.92157034889687 |
| 5-b | 7.906694182348759 | 22.771948777963715 |
| 6-a | 6.315157336730044 | 22.466305309372895 |
| 6-b | 6.174138574511356 | 22.331380641103525 |

Q4. Να βρεθούν οι τρεις μεγαλύτερες (top 3) ώρες αιχμής ανά ημέρα της εβδομάδος, εννοώντας τις ώρες (π.χ., 7-8πμ, 3-4μμ, κλπ) της ημέρας με τον μεγαλύτερο αριθμό επιβατών σε μια κούρσα ταξί. Ο υπολογισμός αφορά όλους τους μήνες.

| | | |
|----------------|--------------|--------------|
| | 1 Worker | 2 Workers |
| Execution Time | 22.46552 sec | 22.40782 sec |

| pickup_day | pickup_hour | avg_passenger_count |
|------------|-------------|---------------------|
| 1 | 0 | 1.5299456507188562 |
| 1 | 1 | 1.527838567375201 |
| 1 | 2 | 1.5080726185191242 |
| 2 | 0 | 1.4679887711672552 |
| 2 | 1 | 1.4442867916810471 |
| 2 | 2 | 1.4231993989051486 |
| 3 | 0 | 1.4200313882151518 |
| 3 | 1 | 1.4175124740006593 |
| 3 | 2 | 1.4104520814693964 |
| 4 | 0 | 1.4012291857176276 |
| 4 | 1 | 1.4088480212656305 |
| 4 | 2 | 1.401127528255003 |
| 5 | 0 | 1.401035624160194 |
| 5 | 1 | 1.402619645711676 |
| 5 | 23 | 1.4053823152498932 |

| | | |
|---|----|--------------------|
| 6 | 2 | 1.4230581143524386 |
| 6 | 22 | 1.444813976205668 |
| 6 | 23 | 1.475576918073731 |
| 7 | 0 | 1.4993154284898547 |
| 7 | 22 | 1.5068176194011382 |
| 7 | 23 | 1.522606766277207 |

Q5. Να βρεθούν οι κορυφαίες πέντε (top 5) ημέρες ανά μήνα στις οποίες οι κούρσες είχαν το μεγαλύτερο ποσοστό σε tip. Για παράδειγμα, εάν η κούρσα κόστισε 10\$ fare_amount και το tip ήταν 5\$, το ποσοστό είναι 50%:

| | 1 Worker | 2 Workers |
|----------------|--------------|--------------|
| Execution Time | 32.00583 sec | 23.42906 sec |

| pickup_month | pickup_day | tip_percentage_avg |
|--------------|------------|--------------------|
| 1 | 1 | 29.063019398121114 |
| 1 | 9 | 45.78674775487207 |
| 1 | 16 | 23.377299918220096 |
| 1 | 29 | 24.059518454370057 |
| 1 | 31 | 43.93563580770273 |
| 2 | 9 | 23.904535643412483 |
| 2 | 10 | 23.33961589934868 |
| 2 | 13 | 24.572068389402546 |
| 2 | 21 | 25.981657452766274 |
| 2 | 27 | 23.3006799515465 |
| 3 | 5 | 22.55546137249565 |
| 3 | 12 | 22.100859110808635 |
| 3 | 18 | 29.671341612659685 |

| | | |
|---|----|--------------------|
| 3 | 21 | 27.57992602492248 |
| 3 | 26 | 22.70884595372165 |
| 4 | 2 | 31.175092883998943 |
| 4 | 3 | 24.463727704753754 |
| 4 | 12 | 48.36884410450343 |
| 4 | 21 | 30.44861250236278 |
| 4 | 30 | 21.99676965994668 |
| 5 | 6 | 21.832006161884486 |
| 5 | 12 | 32.402658973198044 |
| 5 | 15 | 22.05244524700949 |
| 5 | 16 | 23.659110789279985 |
| 5 | 20 | 26.034036090366385 |
| 6 | 10 | 27.397637812780694 |
| 6 | 13 | 38.45136993724611 |
| 6 | 16 | 25.534975757875227 |
| 6 | 20 | 24.242914593519107 |
| 6 | 25 | 32.91307329265353 |

Συνολική απεικόνιση των παραπάνω:

