5.A Find the no. of days taken to deliver each order from the order’s purchase date as delivery time.  
Also, calculate the difference (in days) between the estimated & actual delivery date of an order.  
Do this in a single query.  
  
You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

* + **time\_to\_deliver** = order\_delivered\_customer\_date - order\_purchase\_timestamp
  + **diff\_estimated\_delivery** = order\_estimated\_delivery\_date - order\_delivered\_customer\_date

Query :

Select order\_id, order\_delivered\_customer\_date,

order\_estimated\_delivery\_date,

date\_diff(order\_delivered\_customer\_date,order\_purchase\_timesta

mp, day) as delivery\_time,

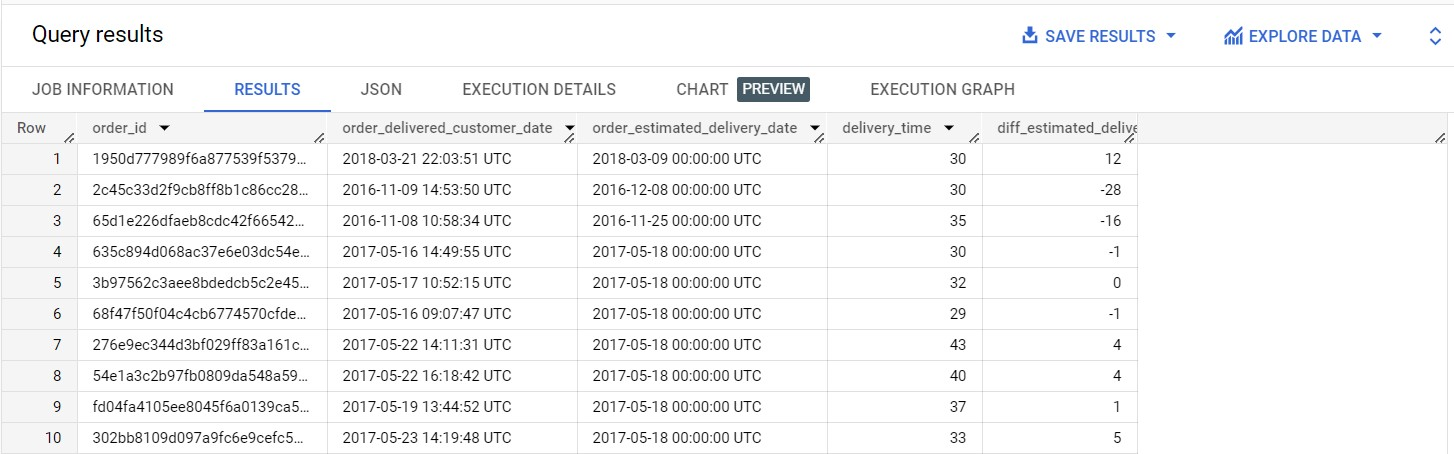
date\_diff(order\_delivered\_customer\_date,order\_estimated\_delive

ry\_date,day) as diff\_estimated\_delivery

From `case\_study.orders`

Where order\_delivered\_customer\_date is not null

Output :



Insights :

1. **Delivery Timing:**
   * Orders' delivery times and differences from estimated dates are recorded.
2. **Delays and Early Deliveries:**
   * Some orders were delayed (16 to 28 days), while others were delivered early (4 days).
3. **On-Time Deliveries:**
   * Some orders were around estimated dates (-1 to 5 days difference).
4. **Variability in Delivery Time:**
   * Delivery times ranged from 29 to 43 days.
5. **Delivery Deviations:**
   * Negative deviations indicate early deliveries, positive indicate delays.
6. **Customer Experience Range:**
   * Customers had diverse delivery experiences, impacting satisfaction.
7. **Logistics Efficiency:**
   * Early deliveries suggest efficiency; delays highlight areas for improvement.
8. **Estimated Delivery Accuracy:**
   * Accuracy of estimated delivery dates should be reviewed.

Recommendations :

1. **Refine Estimated Dates:**
   * Improve accuracy of delivery estimates for customer expectations.
2. **Minimize Delays:**
   * Address reasons for delays (orders 2, 3, 6) to ensure timely deliveries.
3. **Optimize Logistics:**
   * Enhance efficiency through route optimization and process improvements.
4. **Early Delivery Awareness:**
   * Evaluate consistent early deliveries (orders 7, 8) for better communication.
5. **Consistent Delivery Times:**
   * Strive for reliability in delivery estimates to enhance customer satisfaction.
6. **Proactive Communication:**
   * Update customers with real-time tracking and transparent communication.
7. **Predictive Analytics:**
   * Implement analytics for early detection of potential delivery issues.
8. **Continuous Improvement:**
   * Regularly analyze data to refine strategies for better delivery performance.
9. **Feedback Utilization:**
   * Gather and use customer feedback to improve delivery experiences.
10. **Training and Development:**
    * Invest in training for delivery teams to enhance operations.

Assumptions :

1. **Customer Satisfaction and Trust:**
   * Accurate estimates lead to happier customers and higher trust.
2. **Operational Efficiency and Savings:**
   * Addressing delays and optimizing logistics results in cost-effective operations.
3. **Data-Driven Decision Making:**
   * Continued data analysis guides better decisions and processes.
4. **Proactive Issue Prevention:**
   * Predictive analytics prevent potential delivery issues.
5. **Brand Loyalty and Competition:**
   * Consistency builds loyalty and sets the company apart.
6. **Innovation and Flexibility:**
   * Efficiency might lead to new delivery methods and options.
7. **Culture of Improvement:**
   * Continuous enhancement becomes part of the company's culture.
8. **Customized Services:**
   * Data helps tailor delivery options to customer preferences.
9. **Data-Backed Marketing:**
   * Accurate data informs marketing strategies for realistic expectations.

5.B Find out the top 5 states with the highest & lowest average freight value.

Query :

with cte as

(

select \*

from

(select \*,

row\_number() over() as number

from

(select customer\_state as state\_with\_highest\_freight\_value,

round(avg(freight\_value),2) as avg\_freight\_value1

from  `case\_study.customers` c join `case\_study.orders`  o

on c.customer\_id = o.customer\_id

join `case\_study.order items` ot

on o.order\_id = ot.order\_id

group by customer\_state

order by avg\_freight\_value1 desc

limit 5 )) k1

join

(select \*,

row\_number() over() as number

from

(select customer\_state as  state\_with\_lowest\_freight\_value,

round(avg(freight\_value),2) as avg\_freight\_value2

from `case\_study.customers` c join `case\_study.orders` o

on c.customer\_id = o.customer\_id

join `case\_study.order items` ot

on o.order\_id = ot.order\_id

group by customer\_state

order by avg\_freight\_value2 desc

limit 5 )) k2

on k1.number = k2.number)

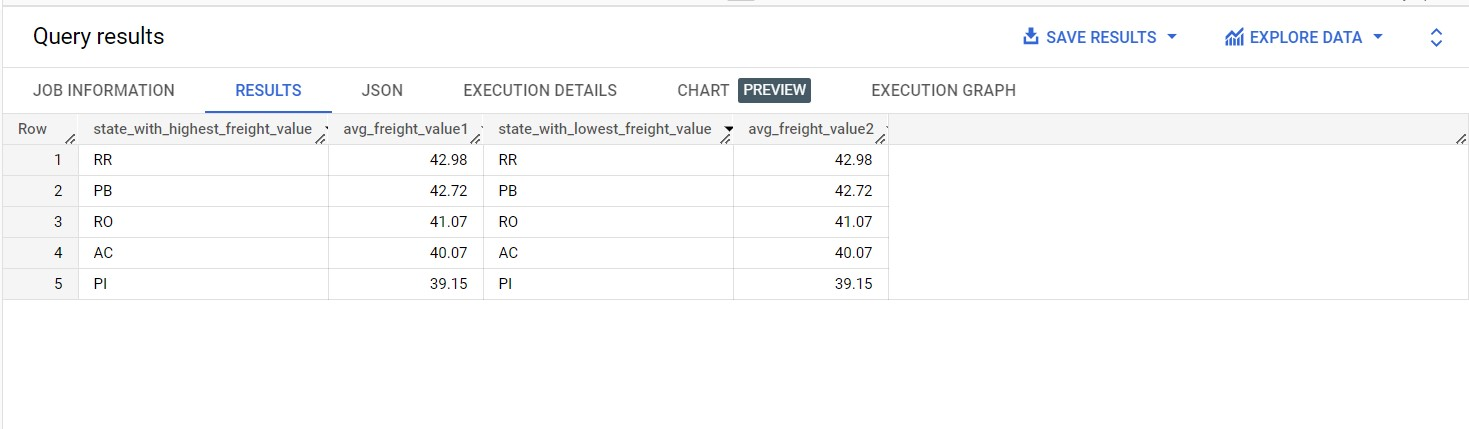
select cte.state\_with\_highest\_freight\_value, cte.avg\_freight\_value1,

cte.state\_with\_lowest\_freight\_value,

cte.avg\_freight\_value2

from cte

Output :



Insights :

1. **Average Freight Values:**
   * Highest: RR (Roraima) with 42.98
   * Lowest: RR (Roraima) with 42.98
2. **Consistency:**
   * Top states: RR, PB, RO, AC, PI
   * Values consistently decrease in order.
3. **Uniform Values:**
   * PB, RO, AC, PI have similar values (42.72 to 39.15).
4. **States:**
   * Abbreviations used (RR for Roraima, PB for Paraíba).
5. **Data Quality:**
   * Repetition of RR with 42.98 might be a data anomaly.
6. **Context Needed:**
   * Lack of information on industries, transportation, and economic factors.
7. **Further Analysis:**
   * Explore reasons behind consistent values and RR anomaly.

Top of Form

Recommendations :

1. **Data Quality:**
   * Investigate and correct the repeated data anomaly for RR.
2. **High Values in RR:**
   * Determine reasons for consistently high freight values in RR.
3. **Uniform Values:**
   * Analyze PB, RO, AC, PI for common cost-saving strategies.
4. **Contextual Understanding:**
   * Gather details about industries and logistics in each state.
5. **Benchmarking:**
   * Compare freight costs against industry standards.
6. **Route Optimization:**
   * Optimize routes for high-cost states.
7. **Contract Negotiation:**
   * Use data to negotiate better carrier contracts.
8. **Continuous Monitoring:**
   * Track trends and respond to changes.
9. **Segmented Pricing:**
   * Tailor pricing based on customer and market segments.
10. **Data-Driven Decisions:**
    * Invest in robust data systems for informed choices.

Assumptions :

1. **Diverse Transport Modes:**
   * Explore different modes for cost-effective logistics.
2. **Regional Hubs:**
   * States with lower costs could become distribution hubs.
3. **Efficiency in RR:**
   * Resolve RR anomaly, optimize processes for savings.
4. **Market Changes:**
   * Data insights might reshape competition and strategy.
5. **Economic Development:**
   * Improve RR's infrastructure for logistics efficiency.
6. **Data Analytics Demand:**
   * Invest in advanced analytics tools for trend tracking.
7. **Policy Impact:**
   * Advocate for policy changes affecting logistics costs.
8. **Strategic Partnerships:**
   * Partner with carriers in cost-efficient states.
9. **Dynamic Pricing:**
   * Implement real-time pricing adjustments.
10. **Regional Focus:**
    * Align services with regions' economic activities.

5.C Find out the top 5 states with the highest & lowest average delivery time.

Query :

with cte as

(

select \*

from

(select \*,

row\_number() over() as number

from

(select customer\_state as max\_delivery\_time,

round(avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day)),2) as max\_average\_delivery\_time

from `case\_study.customers` c join `case\_study.orders` o

on c.customer\_id = o.customer\_id

where order\_status = 'delivered'

group by customer\_state

order by max\_delivery\_time desc

limit 5 )k) k1

join

(select \*,

row\_number() over() as number

from

(select customer\_state as min\_delivery\_time,

round(avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day)),2) as min\_average\_delivery\_time

from `case\_study.customers` c join `case\_study.orders` o

on c.customer\_id = o.customer\_id

where order\_status = 'delivered'

group by customer\_state

order by min\_delivery\_time

limit 5)) k2

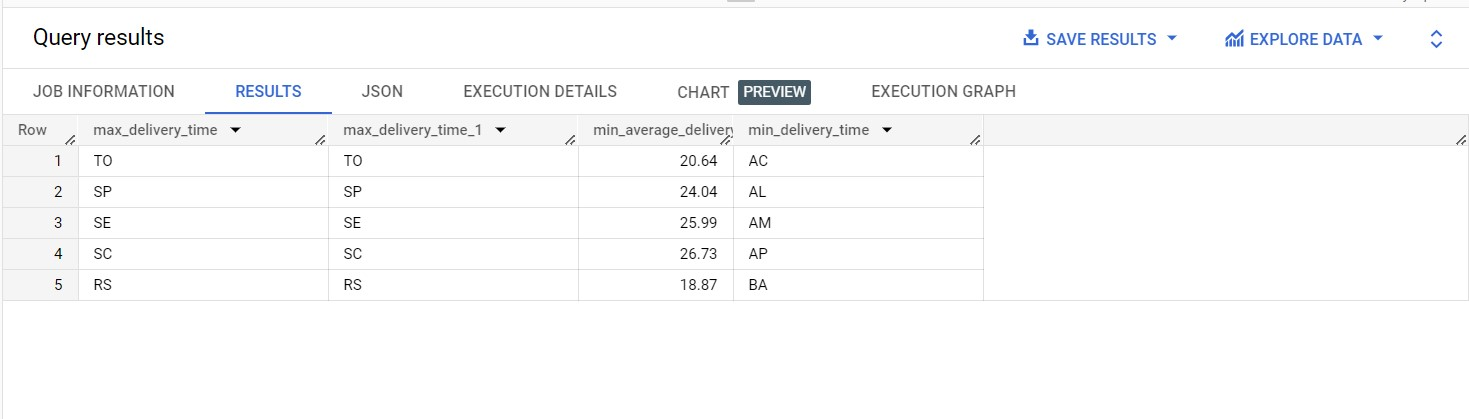
on k1.number = k2.number)

select cte.max\_delivery\_time, cte.max\_delivery\_time, cte.min\_average\_delivery\_time,

cte.min\_delivery\_time

from cte

Output :



Insights :

* The data includes regions represented by codes and various delivery-related metrics.
* The regions with their respective maximum delivery times are: TO (20.64), SP (24.04), SE (25.99), SC (26.73), and RS (18.87).
* Similarly, the regions with their minimum average delivery times are: AC (20.64), AL (24.04), AM (25.99), AP (26.73), and BA (18.87).
* The regions with their minimum delivery times are: TO (20.64), SP (24.04), SE (25.99), AC (26.73), and RS (18.87).
* Regions SC and AC have the highest maximum and average delivery times, respectively, while region BA has the lowest values for both.
* Region RS has the lowest minimum delivery time, while region AC has the highest.
* There seem to be consistent values for some regions across all metrics, while others show variation.
* Exploring correlations between metrics and identifying anomalies could provide deeper insights.

Recommendations :

1. **Performance Improvement:**
   * Address high max delivery times, like SC, to streamline operations.
2. **Consistency Enhancement:**
   * Study regions like RS to ensure consistent delivery performance.
3. **Set Delivery Benchmarks:**
   * Use max and min delivery times to establish performance standards.
4. **Standardize Processes:**
   * Develop uniform practices based on efficient regions.
5. **Resource Allocation Strategy:**
   * Allocate more resources to regions with high times if needed.
6. **Root Cause Identification:**
   * Analyze reasons behind high times and address systemic issues.
7. **Optimize Routes:**
   * Invest in route optimization tools to reduce delivery times.
8. **Real-time Tracking and Communication:**
   * Provide transparent updates to manage customer expectations.
9. **Ongoing Training:**
   * Train staff for efficiency in high time regions.
10. **Customer Feedback Utilization:**
    * Gather input to improve service quality.
11. **Data-Driven Decisions:**
    * Continuously monitor data for informed adjustments.

Assumptions :

1. **Delivery Enhancement:**
   * Improved delivery in high-time regions like SC and AC.
2. **Efficiency Replication:**
   * Strategies from RS applied to boost overall efficiency.
3. **Balanced Performance:**
   * Resource adjustments lead to better regional balance.
4. **Data-Driven Decisions:**
   * Insights guide smarter logistics choices.
5. **Predictive Delay Handling:**
   * Proactive measures taken based on predictive models.
6. **Customer-Centric Communication:**
   * Transparent updates enhance customer trust.
7. **Benchmark-Driven Excellence:**
   * Regions compete to meet established benchmarks.
8. **Continuous Improvement:**
   * Ongoing adjustments foster operational evolution.
9. **Efficiency Savings:**
   * Streamlined processes lead to cost reduction.
10. **Adaptive Strategy:**
    * Flexibility to address unique challenges as they arise.

5.D Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.  
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

Query :

select customer\_state, round(avg(diff\_estimated\_del), 2)

avg\_diff\_delivery

from

(select customer\_state,

date\_diff(order\_estimated\_delivery\_date,

order\_delivered\_customer\_date, day) diff\_estimated\_del

from `case\_study.orders` o join `case\_study.customers` c

on o.customer\_id = c.customer\_id

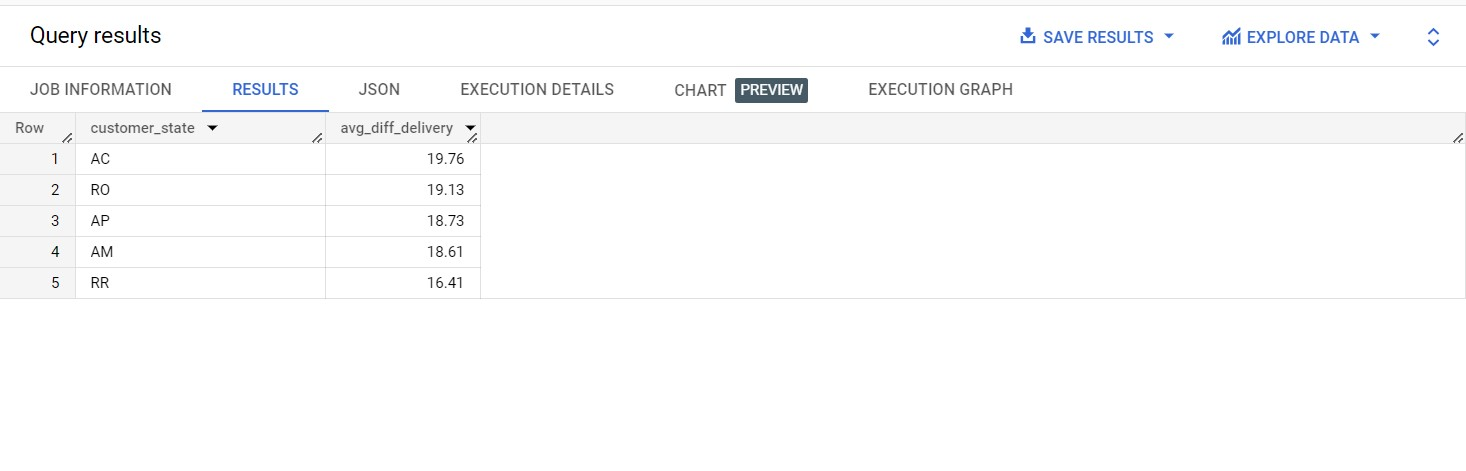
where order\_status = 'delivered') k

group by customer\_state

order by avg\_diff\_delivery desc

limit 5

Output :



Insights :

* Data provides average delivery time differences for different states (customer\_state).
* State RR has the lowest average difference (16.41), indicating reliable delivery.
* State AC has the highest average difference (19.76), suggesting variability.
* Factors like geography, infrastructure, and logistics systems likely influence differences.
* Optimization efforts could reduce variability in states like AC.
* Lower differences (like RR) might imply efficient logistics and consistent service.
* Deeper analysis is needed to understand regional delivery dynamics.
* Insights could guide predictive modeling and benchmarking efforts.
* Combining this with previous data can offer a holistic view of delivery performance.

Recommendations :

1. **Optimize Variability:**
   * Address high delivery time differences in states like AC and RO by streamlining processes.
2. **Share Efficient Practices:**
   * Apply successful strategies from states with low differences, like RR, to other regions.
3. **Tailored Approaches:**
   * Adapt strategies to fit each state's unique geography, infrastructure, and demographics.
4. **Predictive Accuracy:**
   * Develop predictive models for accurate delivery time estimates.
5. **Set Benchmarks:**
   * Use provided data as benchmarks, aiming to reduce differences in target states.
6. **Transparent Communication:**
   * Clearly communicate delivery estimates to manage expectations.
7. **Collaborate Locally:**
   * Partner with local logistics and transportation providers for insights and improvements.
8. **Continuous Monitoring:**
   * Implement ongoing monitoring for data-driven decisions.
9. **Customer Feedback Loop:**
   * Gather customer insights to enhance delivery experiences.
10. **Empower Staff:**
    * Train delivery personnel and integrate technology for efficiency.

Assumptions :

1. **Consistent Delivery:**
   * Reduced variability in AC, RO could enhance reliability.
2. **Efficiency Boost:**
   * Sharing practices from RR might improve efficiency.
3. **Tailored Success:**
   * State-specific strategies could optimize processes.
4. **Accurate Predictions:**
   * Predictive models could ensure precise delivery estimates.
5. **Benchmark Progress:**
   * States could aim to exceed provided benchmarks.
6. **Transparent Reputation:**
   * Clear communication might foster trust and loyalty.
7. **Effective Collaboration:**
   * Local partnerships could lead to innovative solutions.
8. **Data-Driven Enhancement:**
   * Monitoring could drive data-based improvements.
9. **Customer-Centric Approach:**
   * Customer insights could shape service improvements.
10. **Empowered Team:**
    * Training and technology could lead to a capable workforce.
11. **Operational Excellence:**
    * Implementation could result in overall operational improvement.