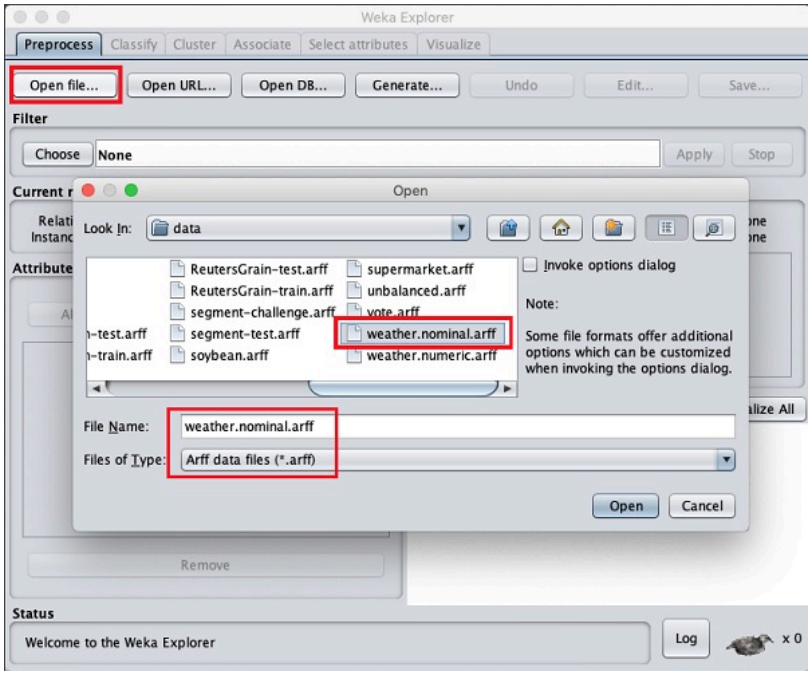
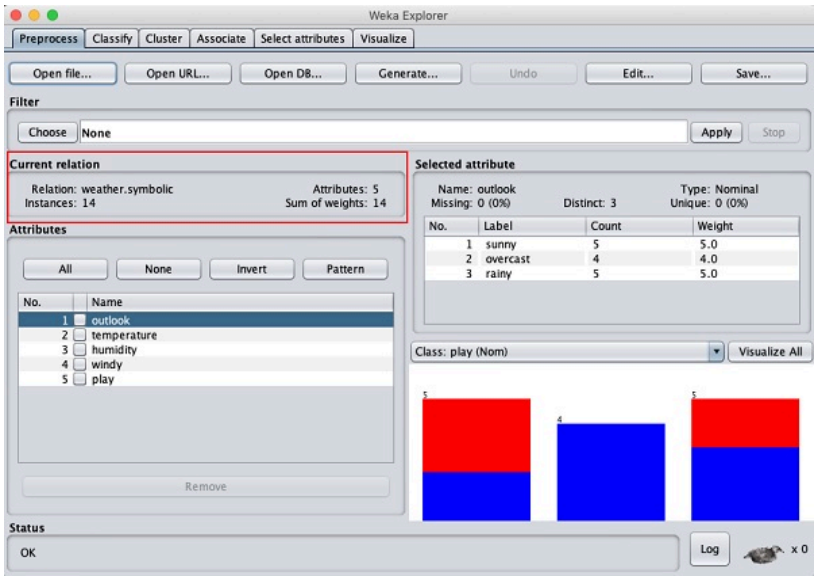
1. Perform data preprocessing tasks using labor data set in WEKA.

The data that is collected from the field contains many unwanted things that leads to wrong analysis. For example, the data may contain null fields, it may contain columns that are irrelevant to the current analysis, and so on. Thus, the data must be preprocessed to meet the requirements of the type of analysis you are seeking. This is the done in the preprocessing module.

Using the **Open file** ... option under the **Preprocess** tag select the **weather-nominal.arff** file.



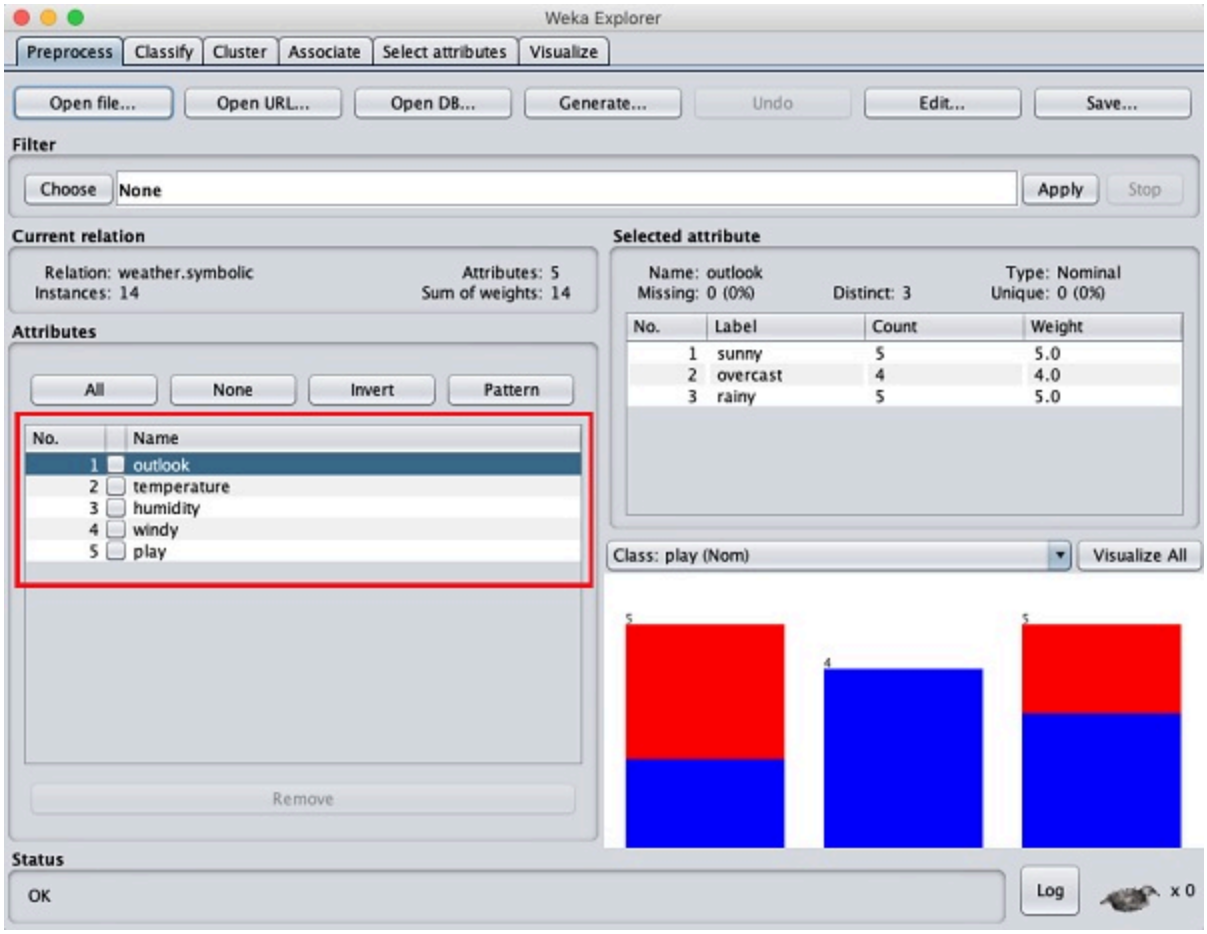


Understanding Data

Let us first look at the highlighted **Current relation** sub window. It shows the name of the database that is currently loaded. You can infer two points from this sub window −

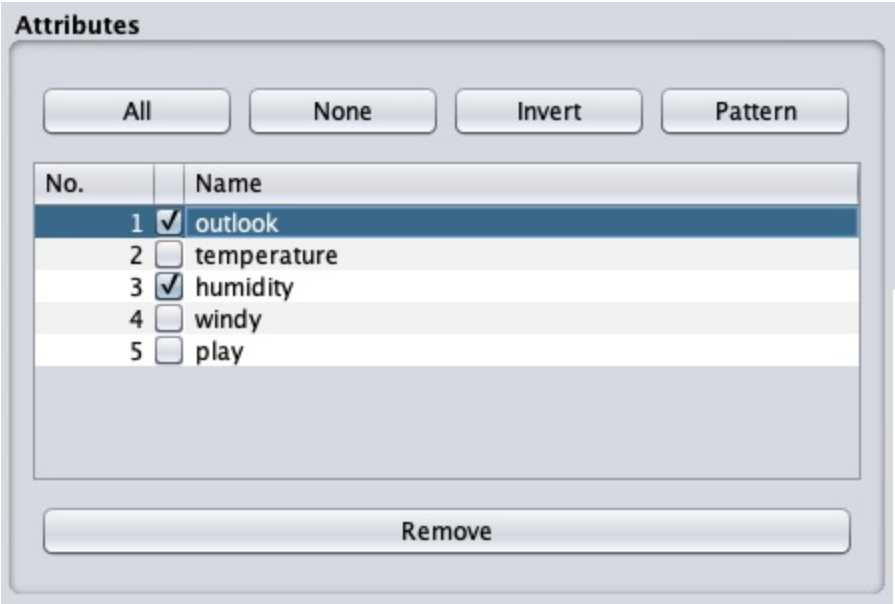
> There are 14 instances - the number of rows in the table.
> The table contains 5 attributes - the fields, which are discussed in the upcoming sections.

On the left side, notice the **Attributes** sub window that displays the various fields in the database.



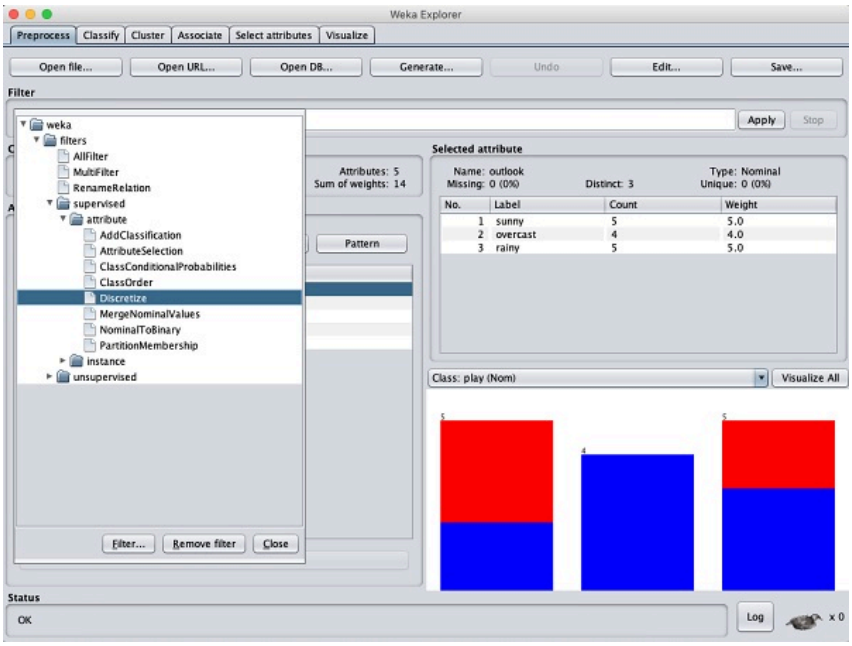
Removing Attributes

Many a time, the data that you want to use for model building comes with many irrelevant fields. For example, the customer database may contain his mobile number which is relevant in analysing his credit rating.



Applying Filters

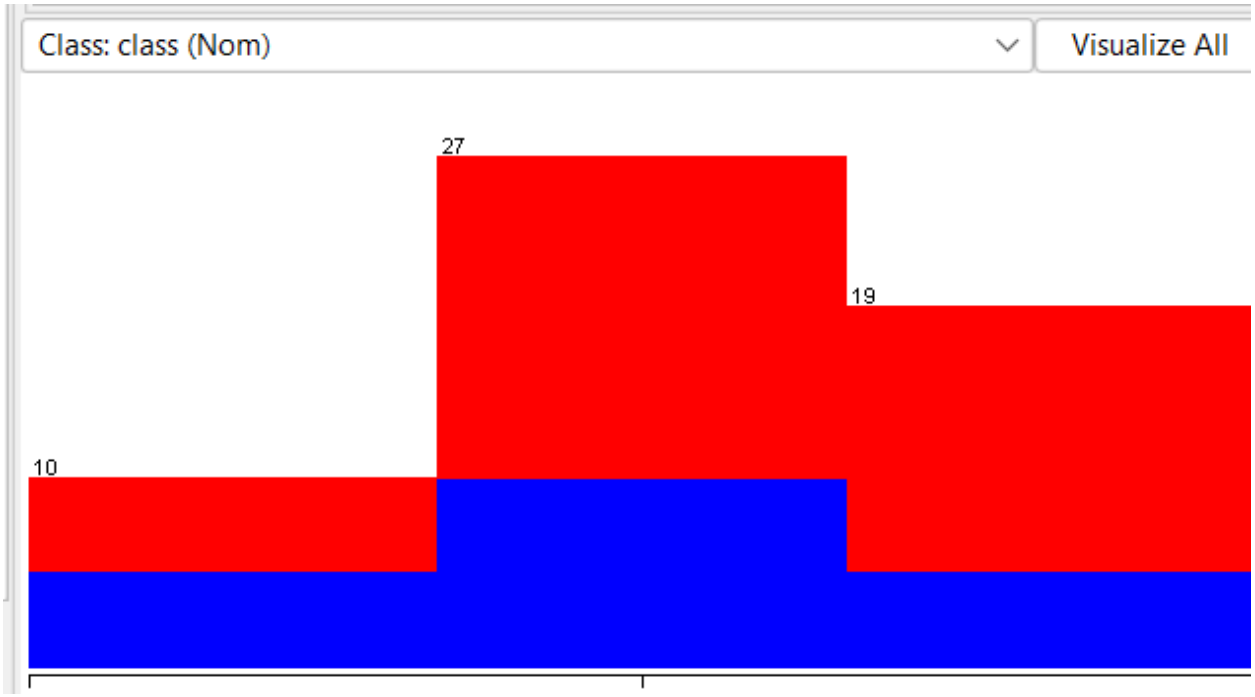
weka→filters→supervised→attribute→Discretize

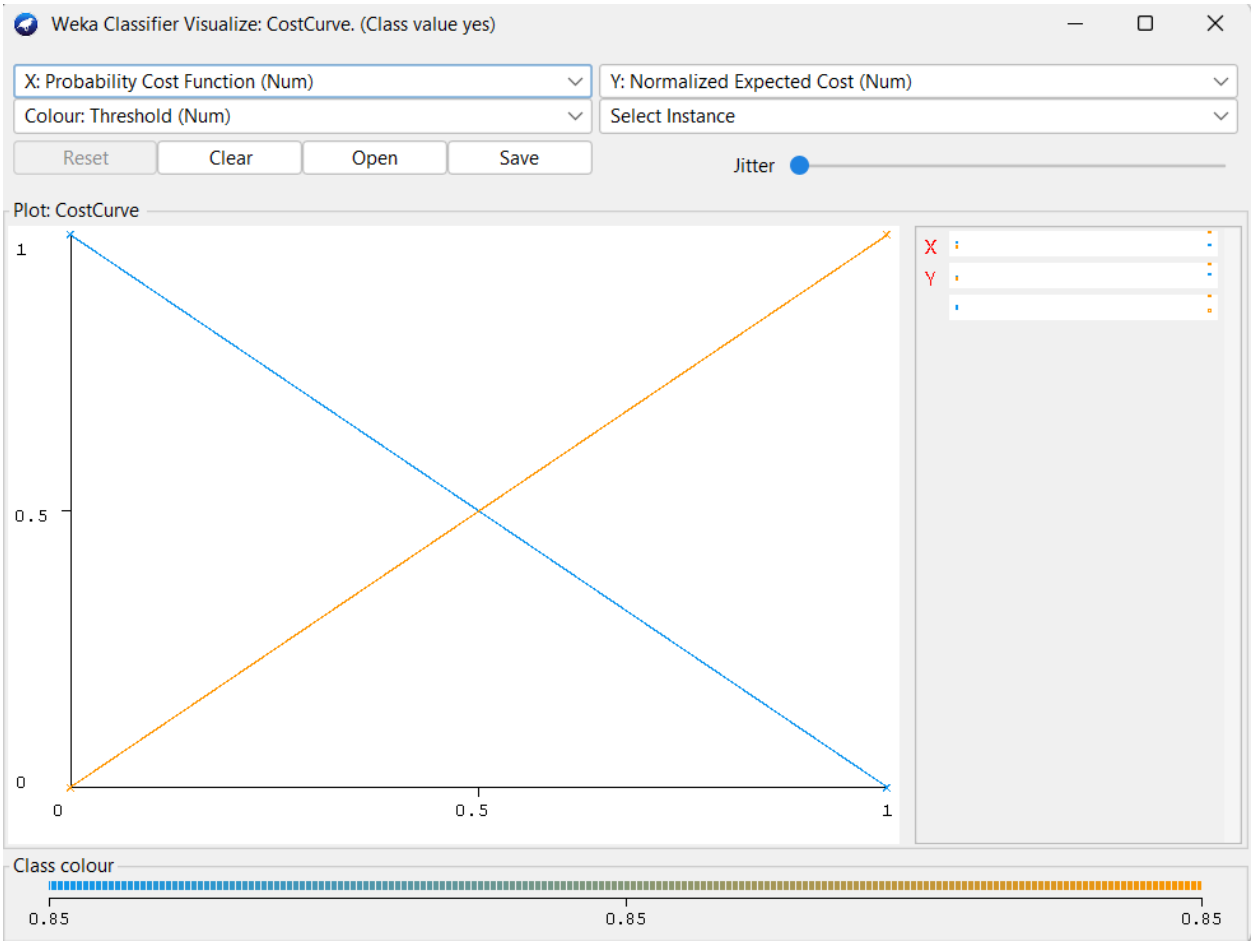


2. Create scatterplots and histograms using visualize option to detect outliers in WEKA.

1. **Open WEKA:** Launch the WEKA application.
2. **Load the Dataset:** Load the labor dataset into WEKA.

3. **Visualize Data:** Go to the "Preprocess" tab, then click on the "Open file" button to load your dataset. Once loaded, select the dataset in the drop-down menu.
4. **Scatterplots:**
    - Click on the "Visualize" tab.
    - In the "Attribute selection" section, select two numeric attributes that you want to plot against each other. You can hold down the Ctrl key (Cmd on Mac) to select multiple attributes.
    - Click on the "Scatterplot" button to generate the scatterplot.
5. **Histograms:**
    - Click on the "Visualize" tab.
    - In the "Attribute selection" section, select the numeric attribute for which you want to create a histogram.
    - Click on the "Histogram" button to generate the histogram.
6. **Detecting Outliers:**
    - Examine the scatterplots and histograms for any unusual patterns or extreme values.
    - Outliers in scatterplots may appear as points that deviate significantly from the overall pattern.
    - In histograms, outliers may manifest as bars with very low or very high frequencies compared to the rest of the data.
7. **Interpreting Results:**
    - In scatterplots, look for points that are far away from the main cluster of points.
    - In histograms, look for bars that are significantly taller or shorter than the others, indicating extreme values.
8. **Handling Outliers:**
    - Depending on the nature of the outliers and the specific requirements of your analysis, you may choose to:
        - Remove the outliers if they are data entry errors or clearly erroneous values.
        - Transform the data using techniques like winsorization or log transformation to reduce the impact of outliers.
        - Keep the outliers if they represent genuine extreme values and are important for your analysis.


Class: class (Nom)
Visualize All
27
19
10

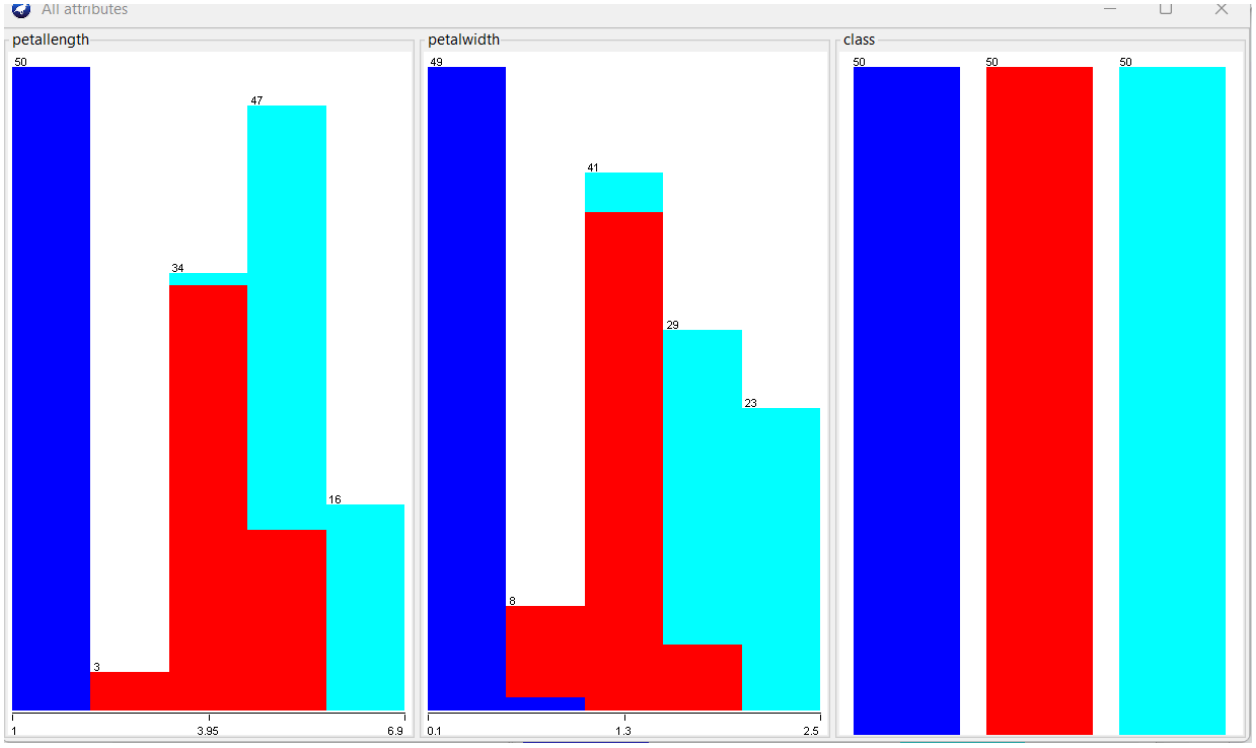

Weka Classifier Visualize: CostCurve. (Class value yes)
X: Probability Cost Function (Num)
Y: Normalized Expected Cost (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: CostCurve
1
0.5
0
0
0.5
1
X
Y
Class colour
0.85
0.85
0.85

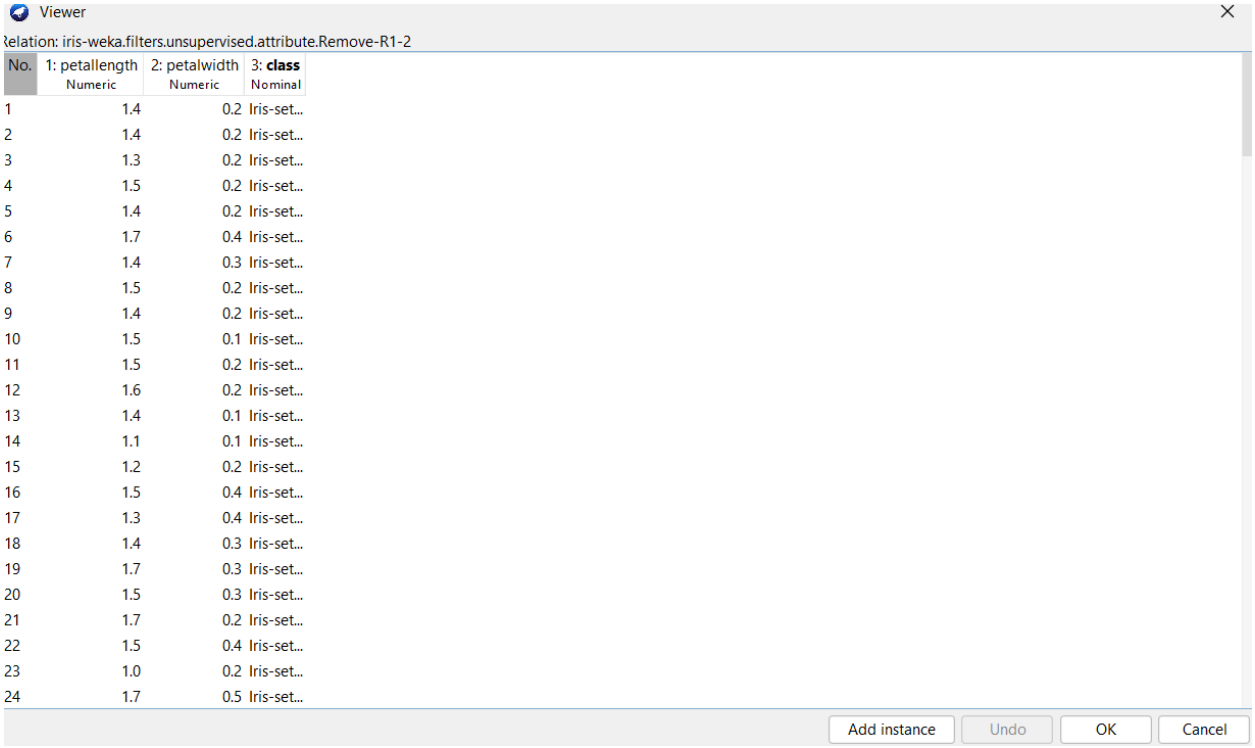
3. Demonstrate how to identify and address conflicting data when merging multiple datasets.

1. **Load Datasets:**
    - Open WEKA.
    - Load the first dataset by clicking on the "Open file" button and selecting the dataset file.
    - Repeat the process to load each additional dataset.
2. **Identify Overlapping Attributes:**
    - Explore each dataset to identify attributes that are common across multiple datasets. You can view the attributes in the "Attributes" tab or by clicking on the dataset name in the "Preprocess" tab.
3. **Examine Attribute Values:**
    - Click on each dataset name in the "Preprocess" tab to switch between datasets.
    - Use the "Visualize" tab to inspect the values of overlapping attributes. You can generate histograms or summary statistics to understand the distribution of attribute values.
    - Pay close attention to any discrepancies or differences in values between datasets.
4. **Resolve Conflicts:**
    - Identify conflicting values in overlapping attributes. Look for inconsistencies such as different data formats, scales, or categories.
    - Decide on a strategy to resolve conflicts based on the nature of the data and your domain knowledge.
    - Apply manual inspection, data transformation, imputation, or data fusion techniques to address conflicting data.
5. **Merge Datasets:**
    - Once conflicts are resolved, merge the datasets using WEKA's functionality for merging or joining datasets.
    - In the "Preprocess" tab, navigate to the "Merge" panel.
    - Choose the appropriate merging method (e.g., appending datasets, joining datasets based on a common key attribute).
    - Configure the merging options as needed and apply the merge operation.
6. **Validate Merged Dataset:**

- After merging the datasets, validate the integrity of the merged dataset.
- Inspect the merged dataset in the "Explorer" tab to ensure that all attributes and instances are included.
- Check for consistency and completeness in the merged dataset, ensuring that conflicts are resolved satisfactorily.

7. **Save Merged Dataset:**
    - Once validated, save the merged dataset in a suitable format for further analysis.
    - Click on the "Save" button in the "Preprocess" tab to save the merged dataset to a file.





4. Implement data preprocessing using ARFF format, CSV format and C4.5 format in WEKA tool.

1.Data Preprocessing with ARFF Format:

a. Load Data:
- Open WEKA.
- Go to the "Preprocess" tab.
- Click on the "Open file" button and select your ARFF file.

b. Explore Data:
- Once the ARFF file is loaded, explore the dataset using the "Attributes" tab to view the attributes and their types.
- Use the "Visualize" tab to generate visualizations like histograms and scatterplots for data exploration.

c. Preprocessing Steps:
- Handle missing values using filters like `ReplaceMissingValues`.

- Convert categorical attributes to numeric using filters like `NominalToBinary` or `NominalToString`.
- Normalize or standardize numeric attributes if needed using filters like `Normalize` or `Standardize`.
- Optionally, perform feature selection using filters like `AttributeSelection`.

d. Save Preprocessed Data:
- Once preprocessing is done, save the preprocessed dataset in ARFF format by clicking on the "Save" button and choosing the ARFF format.

2. Data Preprocessing with CSV Format:

a. Load Data:
- Go to the "Preprocess" tab.
- Click on the "Open file" button and select your CSV file.

b. Explore Data:
- Explore the dataset as described above for the ARFF format.

c. Preprocessing Steps:
- Since WEKA doesn't directly support CSV files for preprocessing, you'll need to convert the CSV file to ARFF format first.
- After loading the CSV file, click on the "Save" button and choose the ARFF format to save the file in ARFF format.
- Proceed with preprocessing steps as described for the ARFF format.

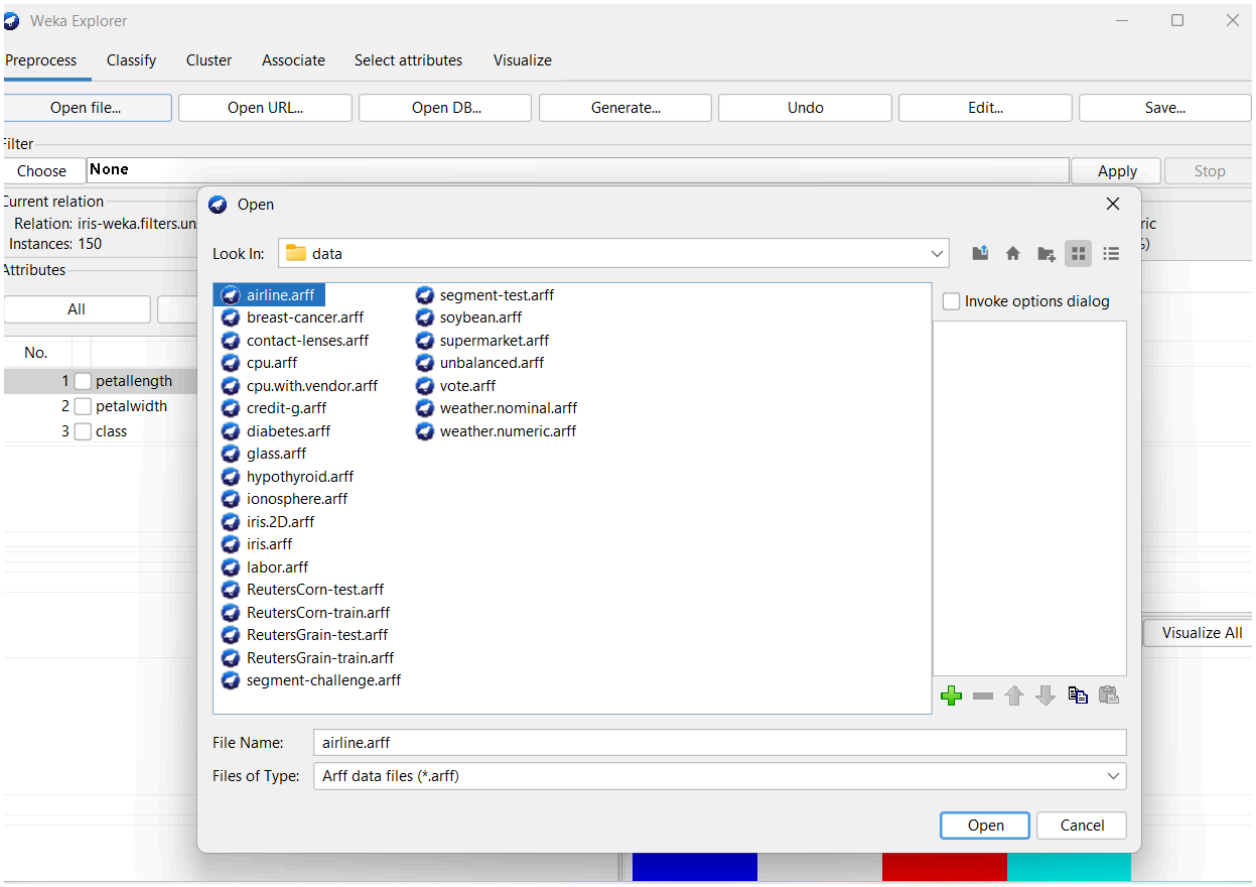
3. Data Preprocessing with C4.5 Format:

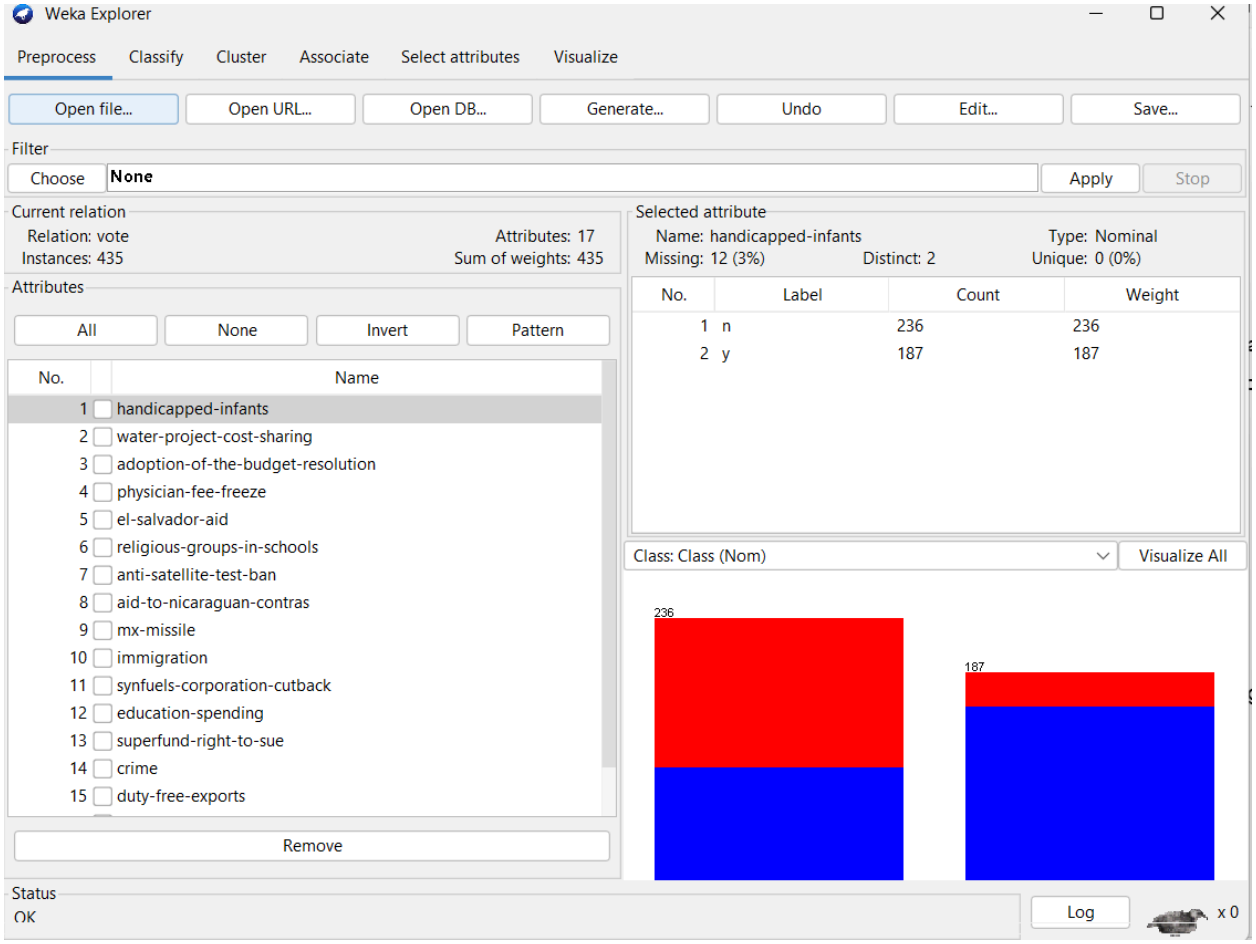
a. Load Data:
- Go to the "Preprocess" tab.
- Click on the "Open file" button and select your C4.5 file.

b. Explore Data:
- Explore the dataset as described above for the ARFF format.

c. Preprocessing Steps:
- Since C4.5 files are typically used for decision tree models, preprocessing steps are similar to those for ARFF files.
- Follow the preprocessing steps described for the ARFF format.

d. Save Preprocessed Data:
- Save the preprocessed dataset in ARFF format as described above.

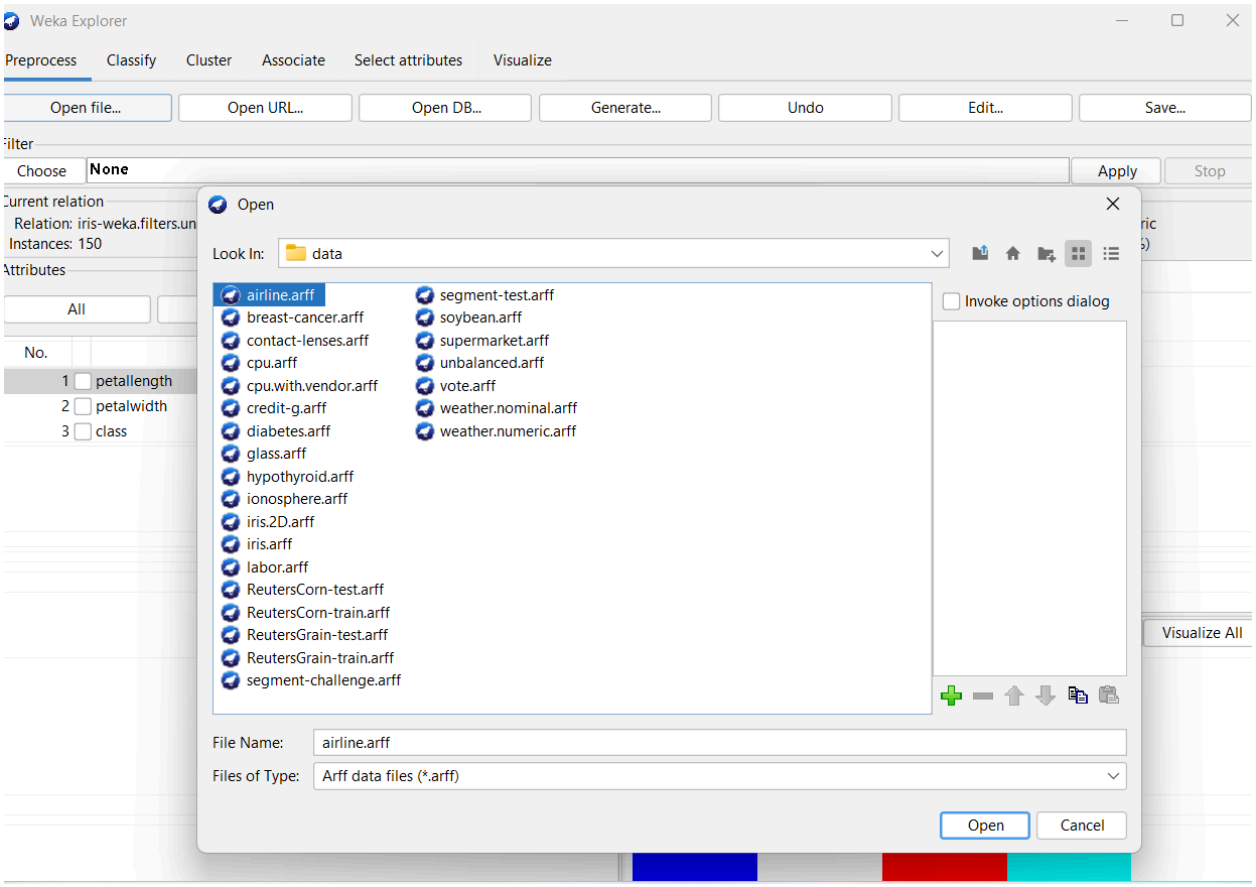

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: iris-weka.filters.un
Instances: 150
Attributes
All
No.
1 petallength
2 petalwidth
3 class
Visualize All
Open
Look In:
data
airline.arff
breast-cancer.arff
contact-lenses.arff
cpu.arff
cpu.with.vendor.arff
credit-g.arff
diabetes.arff
glass.arff
hypothyroid.arff
ionosphere.arff
iris.2D.arff
iris.arff
labor.arff
ReutersCorn-test.arff
ReutersCorn-train.arff
ReutersGrain-test.arff
ReutersGrain-train.arff
segment-challenge.arff
segment-test.arff
soybean.arff
supermarket.arff
unbalanced.arff
vote.arff
weather.nominal.arff
weather.numeric.arff
Invoke options dialog
File Name:
airline.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel

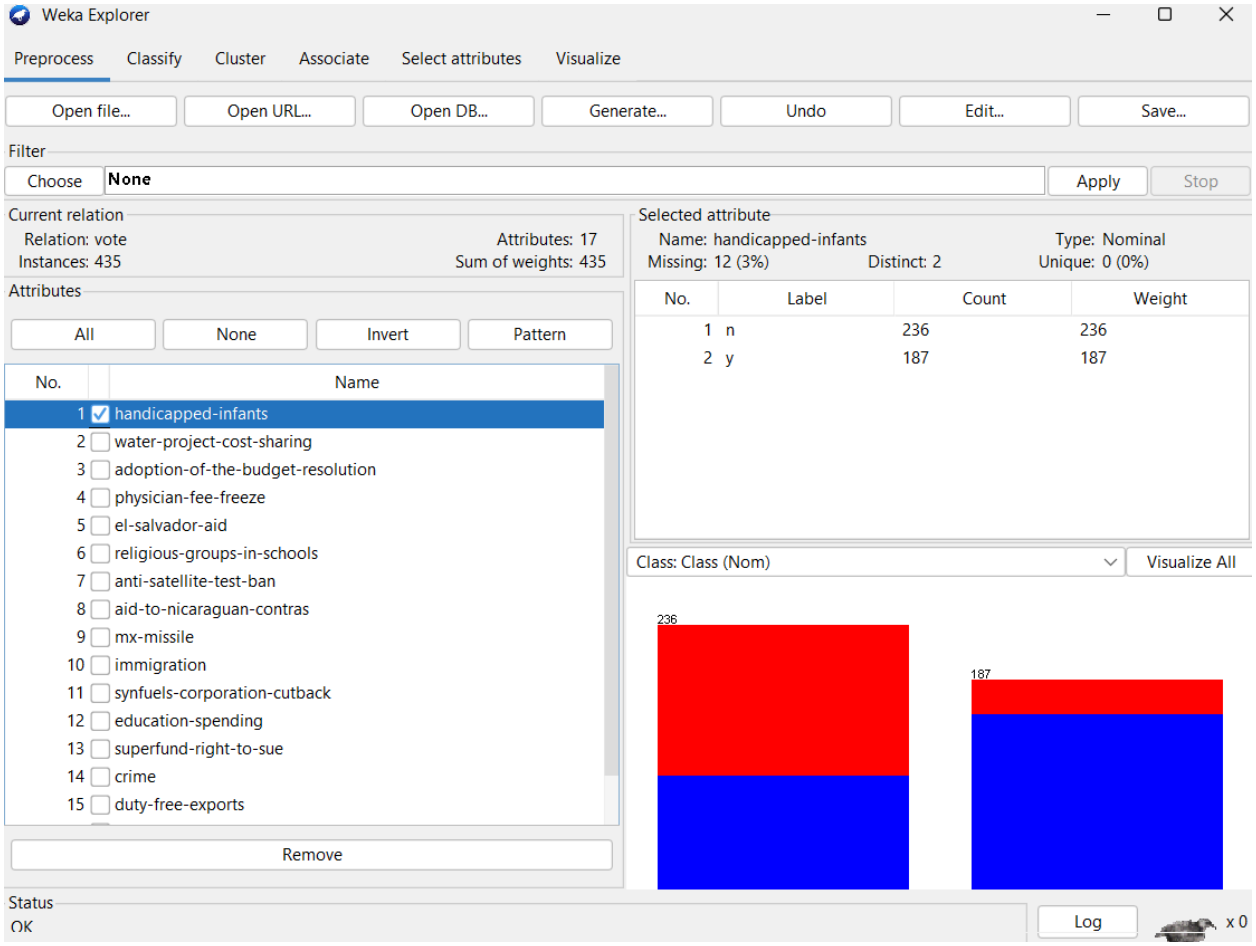


5. Perform data preprocessing tasks using weather database in WEKA. Demonstrate how toremove attributes in the given data base.

1. **Load Weather Database:**
    - Open WEKA.
    - Go to the "Explorer" tab.
    - Click on the "Open file" button.
    - Select the weather database file (e.g., in ARFF format) from your local directory.
    - The dataset will be loaded into WEKA.
2. **Explore Attributes:**
    - In the "Attributes" panel of the "Explorer" tab, you will see a list of all attributes in the loaded dataset.
    - Examine the attributes to identify which ones you want to remove.

3. **Remove Attributes:**
    - Select the attributes you want to remove:
        i. You can click on individual attribute names to select them.
        ii. To select multiple attributes, hold down the Ctrl (Cmd on Mac) key while clicking on the attribute names.
    - Once the attributes are selected, right-click on one of the selected attributes or go to the "Edit" menu at the top.
    - Choose the "Remove" option to delete the selected attributes from the dataset.
4. **Save Preprocessed Dataset (Optional):**
    - If you want to save the dataset with the removed attributes:
        i. Go to the "File" menu.
        ii. Choose the "Save as" option.
        iii. Select the desired format (e.g., ARFF).
        iv. Specify the file name and location to save the preprocessed dataset.
        v. Click "Save" to save the dataset with the removed attributes.
5. **Verify Removed Attributes:**
    - After removing the attributes, you can verify that they have been successfully removed by checking the "Attributes" panel in the "Explorer" tab.
    - The removed attributes should no longer appear in the list of attributes.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: iris-weka.filters.un
Instances: 150
Attributes
All
No.
1 petallength
2 petalwidth
3 class
Visualize All
Open
Look In:
data
airline.arff
breast-cancer.arff
contact-lenses.arff
cpu.arff
cpu.with.vendor.arff
credit-g.arff
diabetes.arff
glass.arff
hypothyroid.arff
ionosphere.arff
iris.2D.arff
iris.arff
labor.arff
ReutersCorn-test.arff
ReutersCorn-train.arff
ReutersGrain-test.arff
ReutersGrain-train.arff
segment-challenge.arff
segment-test.arff
soybean.arff
supermarket.arff
unbalanced.arff
vote.arff
weather.nominal.arff
weather.numeric.arff
Invoke options dialog
File Name:
airline.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel

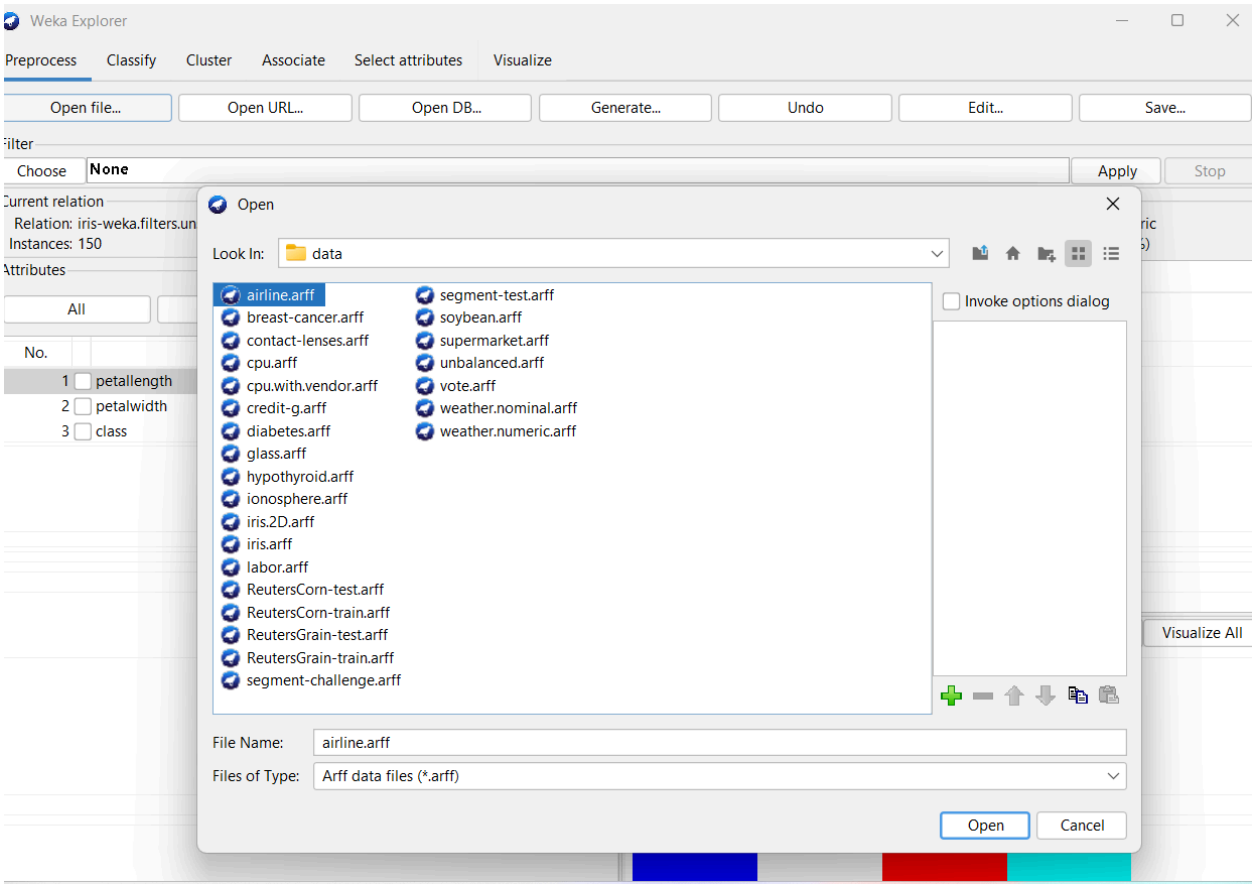


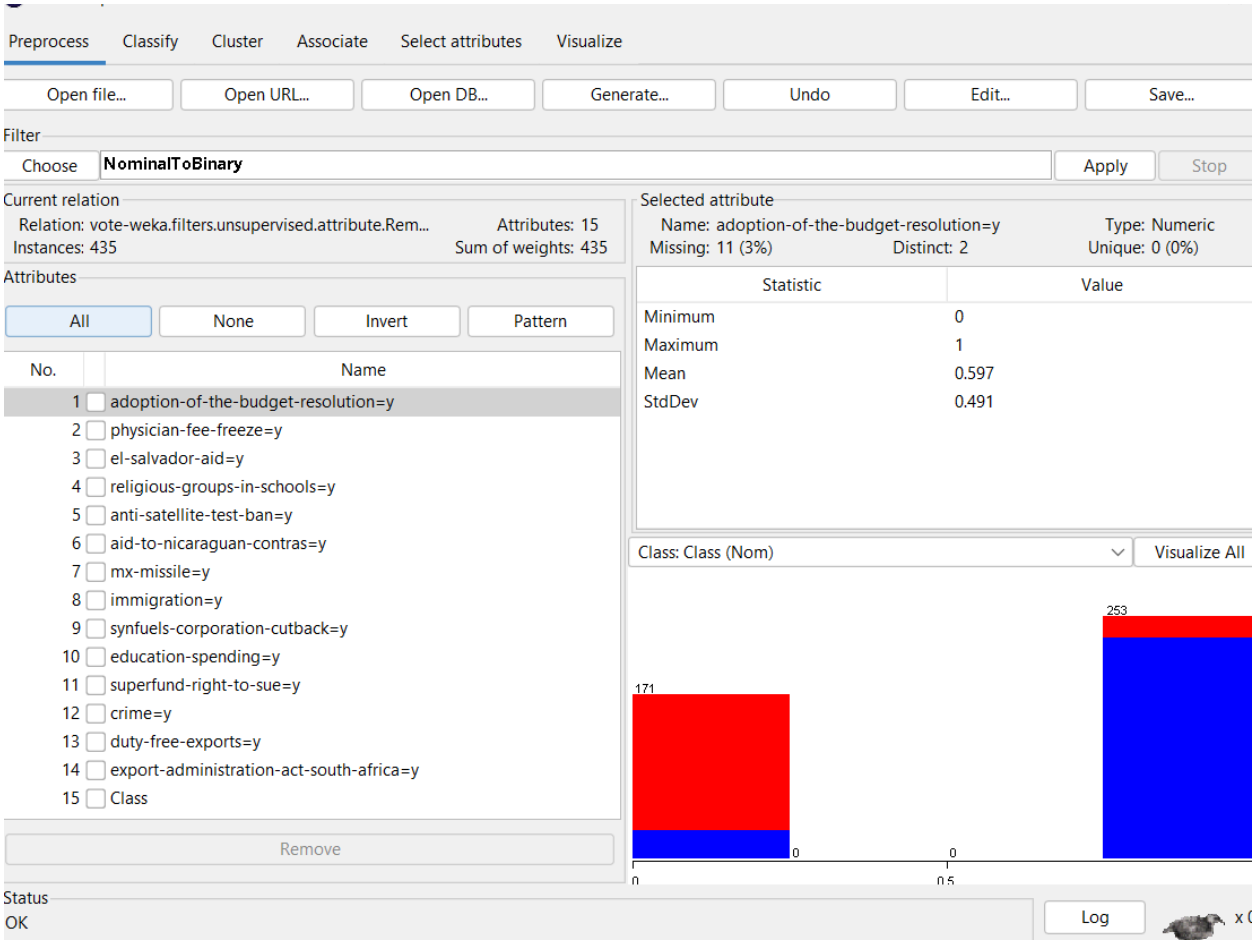
6. Demonstrate the usage of filters in WEKA.

1. **Load Dataset:**
   - Open WEKA.
   - Go to the "Explorer" tab.
   - Click on the "Open file" button.
   - Select the dataset you want to work with (e.g., weather dataset).
2. **Explore Dataset:**
   - Once the dataset is loaded, you can explore its attributes and instances in the "Explorer" tab.
3. **Access Filters:**
   - Navigate to the "Preprocess" tab.
   - Here, you'll find a variety of filters categorized under different categories such as unsupervised, supervised, attribute, instance, etc.
4. **Select Filter:**

- Choose the filter you want to apply from the list. For example, let's say we want to apply the "Remove" filter to remove certain attributes from the dataset.

5. **Configure Filter:**
    - Once you select the filter, a configuration panel will appear where you can adjust the parameters of the filter.
    - For the "Remove" filter, you'll need to specify the indices or names of the attributes you want to remove.
6. **Apply Filter:**
    - After configuring the filter, click on the "Apply" button to apply the filter to the dataset.
    - Depending on the filter, you may also have options to visualize the results or preview the changes before applying them.
7. **Save Preprocessed Dataset:**
    - If you're satisfied with the results of the filter, you can save the preprocessed dataset:
        - Go to the "File" menu.
        - Choose the "Save as" option.
        - Select the desired format (e.g., ARFF).
        - Specify the file name and location to save the preprocessed dataset.
        - Click "Save" to save the dataset with the applied filter.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: iris-weka.filters.un
Instances: 150
Attributes
All
No.
1 petallength
2 petalwidth
3 class
Visualize All
Open
Look In:
data
airline.arff
breast-cancer.arff
contact-lenses.arff
cpu.arff
cpu.with.vendor.arff
credit-g.arff
diabetes.arff
glass.arff
hypothyroid.arff
ionosphere.arff
iris.2D.arff
iris.arff
labor.arff
ReutersCorn-test.arff
ReutersCorn-train.arff
ReutersGrain-test.arff
ReutersGrain-train.arff
segment-challenge.arff
segment-test.arff
soybean.arff
supermarket.arff
unbalanced.arff
vote.arff
weather.nominal.arff
weather.numeric.arff
Invoke options dialog
File Name:
airline.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel



7. Create noise monitoring data set using Apriori algorithm.

1. **Define Attributes:**
    - Attributes include Location (e.g., "Street", "Park", "Office"), Time (e.g., "Morning", "Afternoon", "Evening"), and Noise Level Category (e.g., "Low", "Medium", "High").
2. **Generate Transactions:**
    - Generate transactions representing instances where noise measurements are taken.
    - Each transaction should contain a combination of attributes (location, time, noise level category).
    - Example transactions include:
        - Transaction 1: {Location=Street, Time=Morning, Noise=Low}
        - Transaction 2: {Location=Park, Time=Afternoon, Noise=Medium}
        - Transaction 3: {Location=Office, Time=Evening, Noise=High}

3. **Save Dataset:**
    - Save the dataset in ARFF format, where each transaction is represented as an instance with attributes corresponding to location, time, and noise level category.
4. **Apply Apriori Algorithm:**
    - Load the dataset into WEKA.
    - Use the Apriori algorithm to mine association rules.
    - Configure the algorithm to identify frequent itemsets, representing combinations of attributes, and generate association rules based on these itemsets.
5. **Interpret Results:**
    - Analyze the generated association rules to understand relationships between different attributes.
    - For example, you might discover rules such as "High noise levels are more frequent in the evening in the street area."

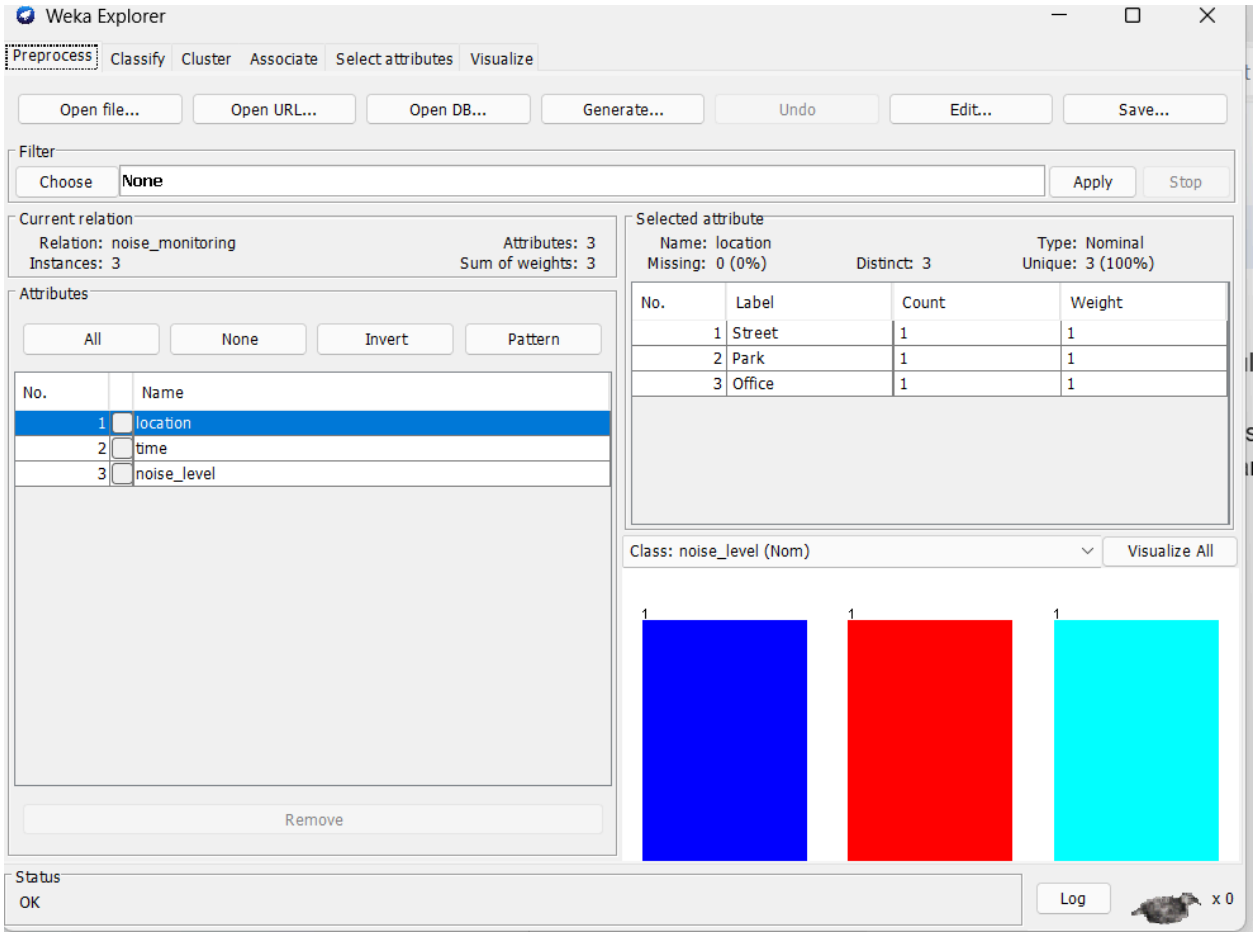
Program 👍

```
@relation noise_monitoring

@attribute location {Street, Park, Office}
@attribute time {Morning, Afternoon, Evening}
@attribute noise_level {Low, Medium, High}

@data
Street, Morning, Low
Park, Afternoon, Medium
Office, Evening, High
```

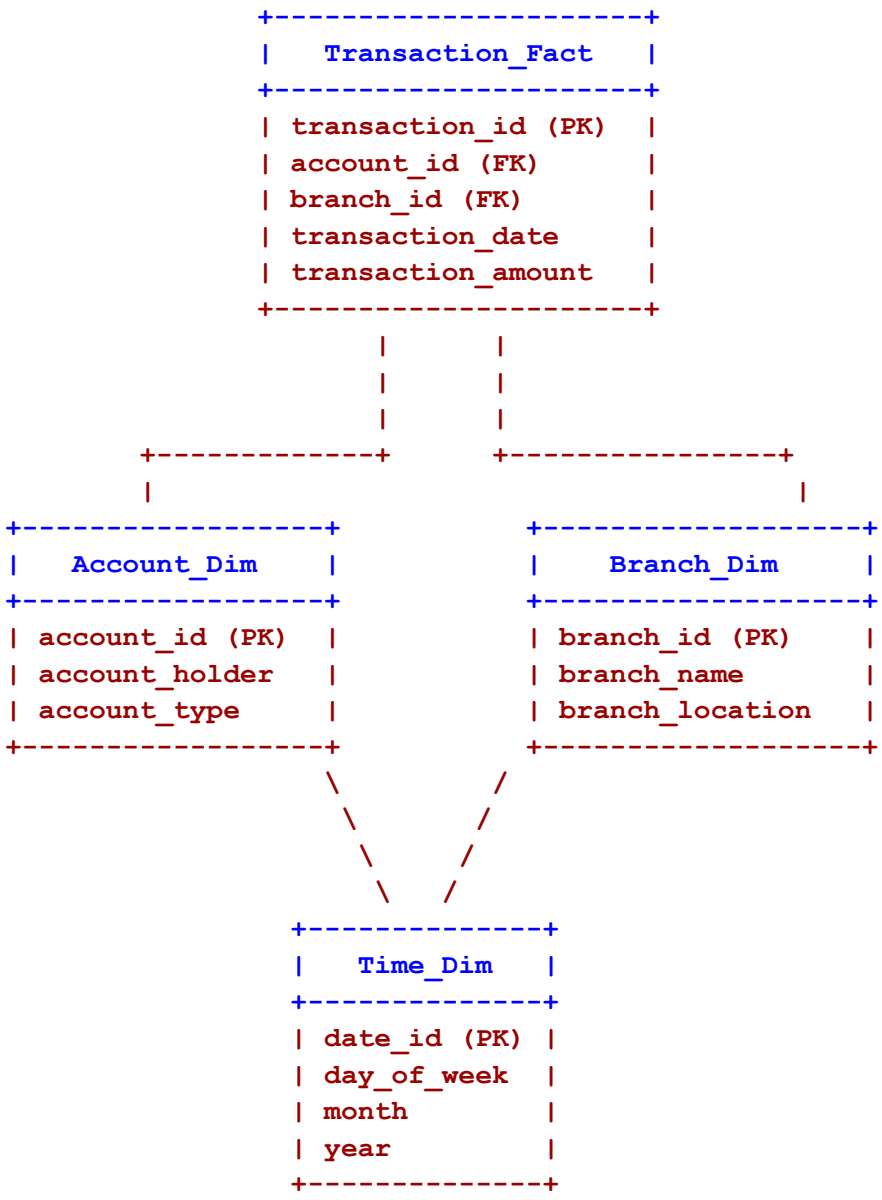
Output:



8. Design multi-dimensional data models such as Star, Snowflake and Fact Constellation schemas for Banking application.

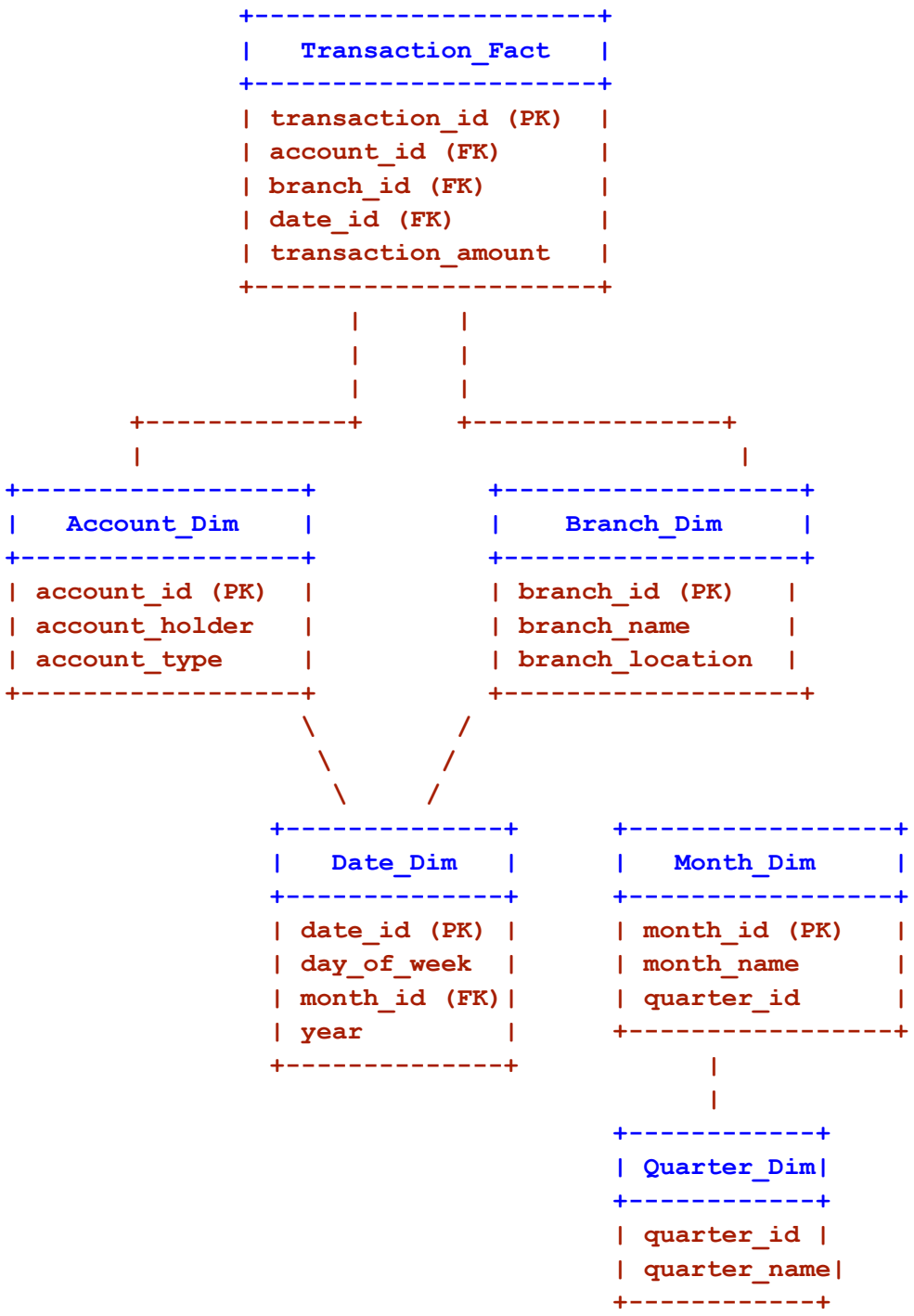
1. **Star Schema:**
    - In a Star Schema, you have a central fact table surrounded by dimension tables.
    - Fact Table: Contains quantitative measures (e.g., transaction amount, balance).
    - Dimension Tables: Contains descriptive attributes (e.g., account holder, branch, time).
    - Example:
        - Fact Table: Transaction_Fact (transaction_id, account_id, branch_id, transaction_date, transaction_amount)

- Dimension Tables: Account_Dim (account_id, account_holder_name, account_type), Branch_Dim (branch_id, branch_name, branch_location), Time_Dim (date_id, day_of_week, month, year)



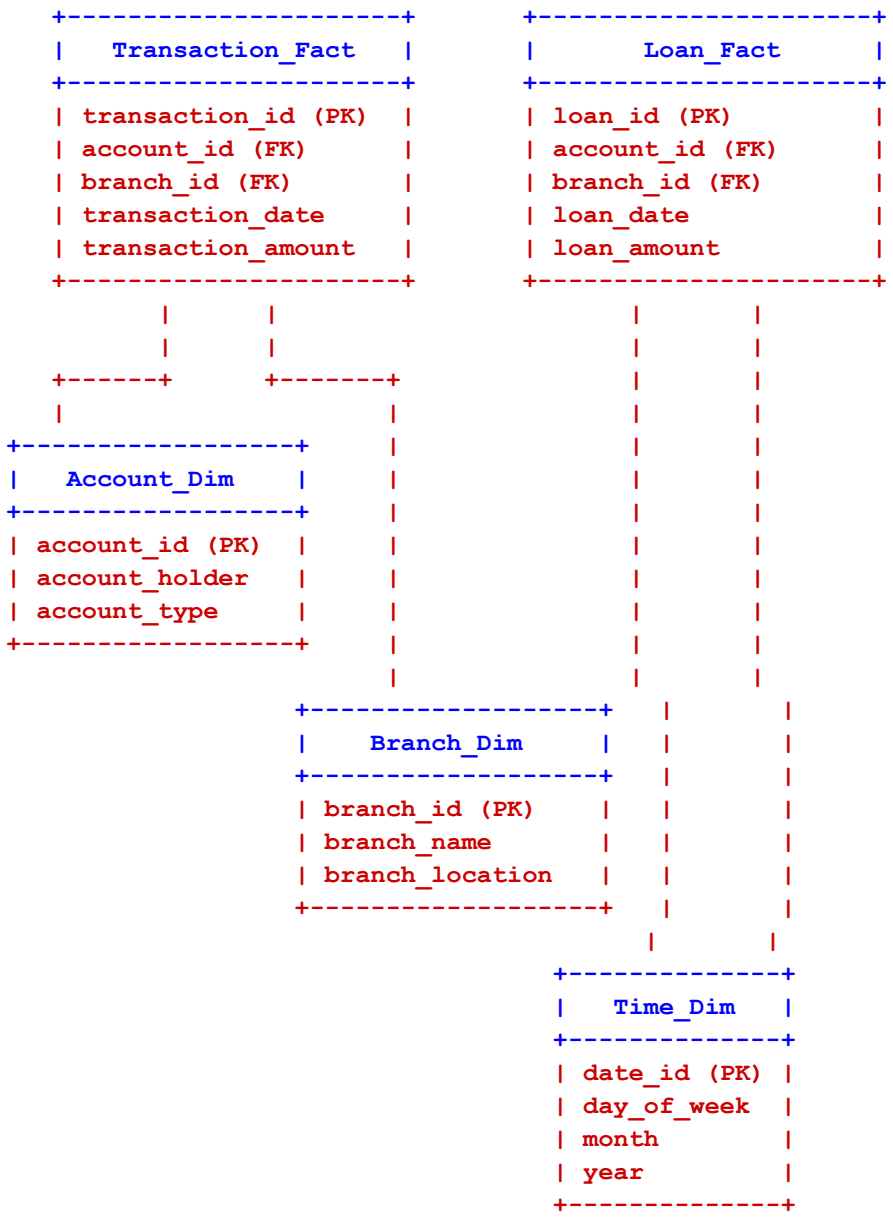
2. **Snowflake Schema:**
   - Snowflake schema is an extension of the Star Schema where dimension tables are normalized into multiple levels.
   - Dimension tables are broken down into smaller tables, resulting in more normalized data.
   - Example:
     - Fact Table: Same as Star Schema

- Dimension Tables: Account_Dim (account_id, account_holder_name, account_type), Branch_Dim (branch_id, branch_name, branch_location), Date_Dim (date_id, day_of_week, month_id), Month_Dim (month_id, month_name, quarter_id), Quarter_Dim (quarter_id, quarter_name)



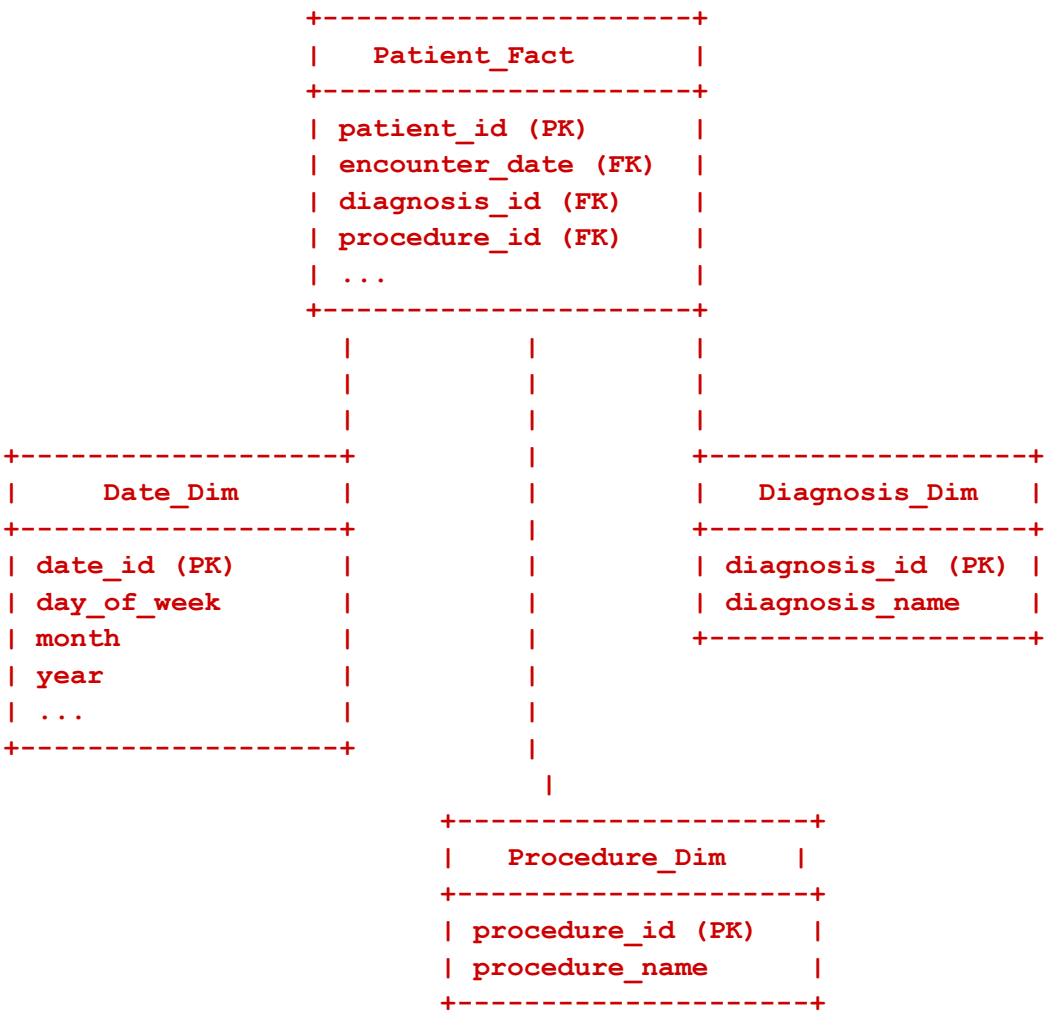
3. **Fact Constellation Schema:**

- Fact Constellation schema consists of multiple fact tables sharing dimension tables.
- Each fact table represents a different business process or aspect of the business.
- Example:
    - Fact Tables: Transaction_Fact (transaction_id, account_id, branch_id, transaction_date, transaction_amount), Loan_Fact (loan_id, account_id, branch_id, loan_date, loan_amount)
    - Dimension Tables: Same as Star Schema

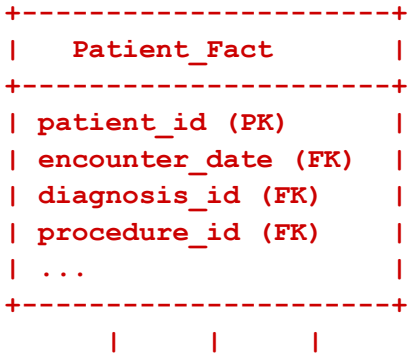


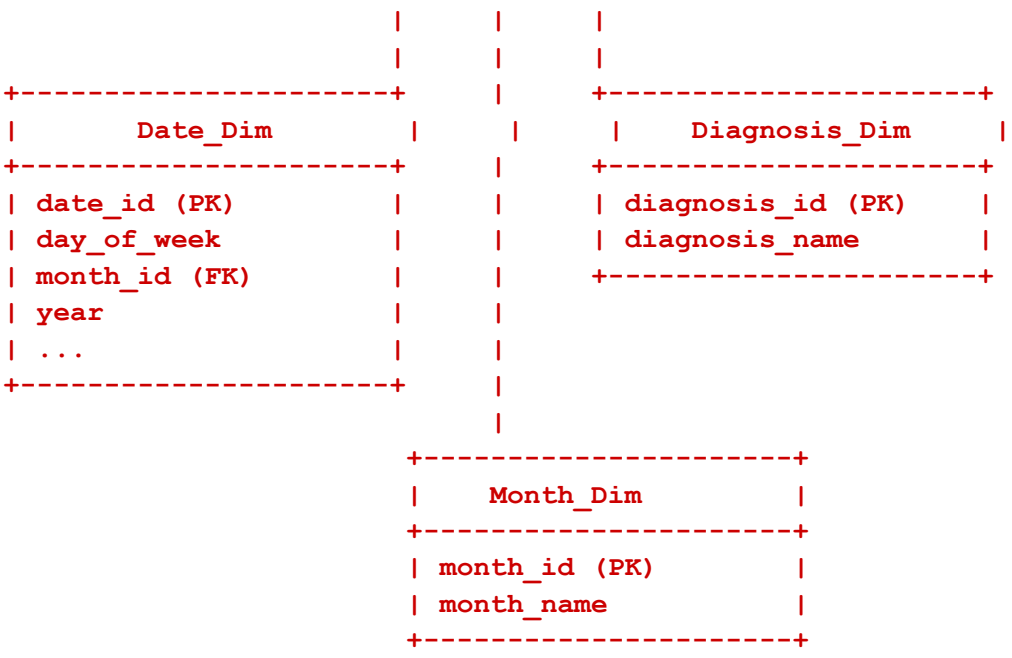
9. Design multi-dimensional data models such as Star, Snowflake and Fact Constellation schemas for Health care application.

1. **Star Schema:**

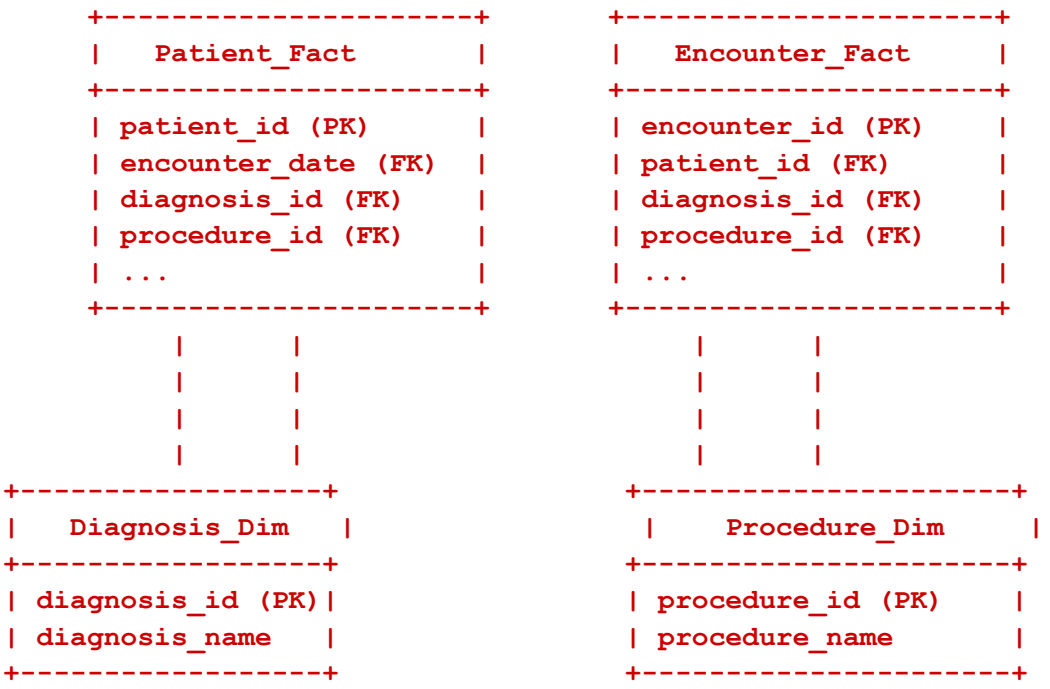


2. **Snowflake Schema:**



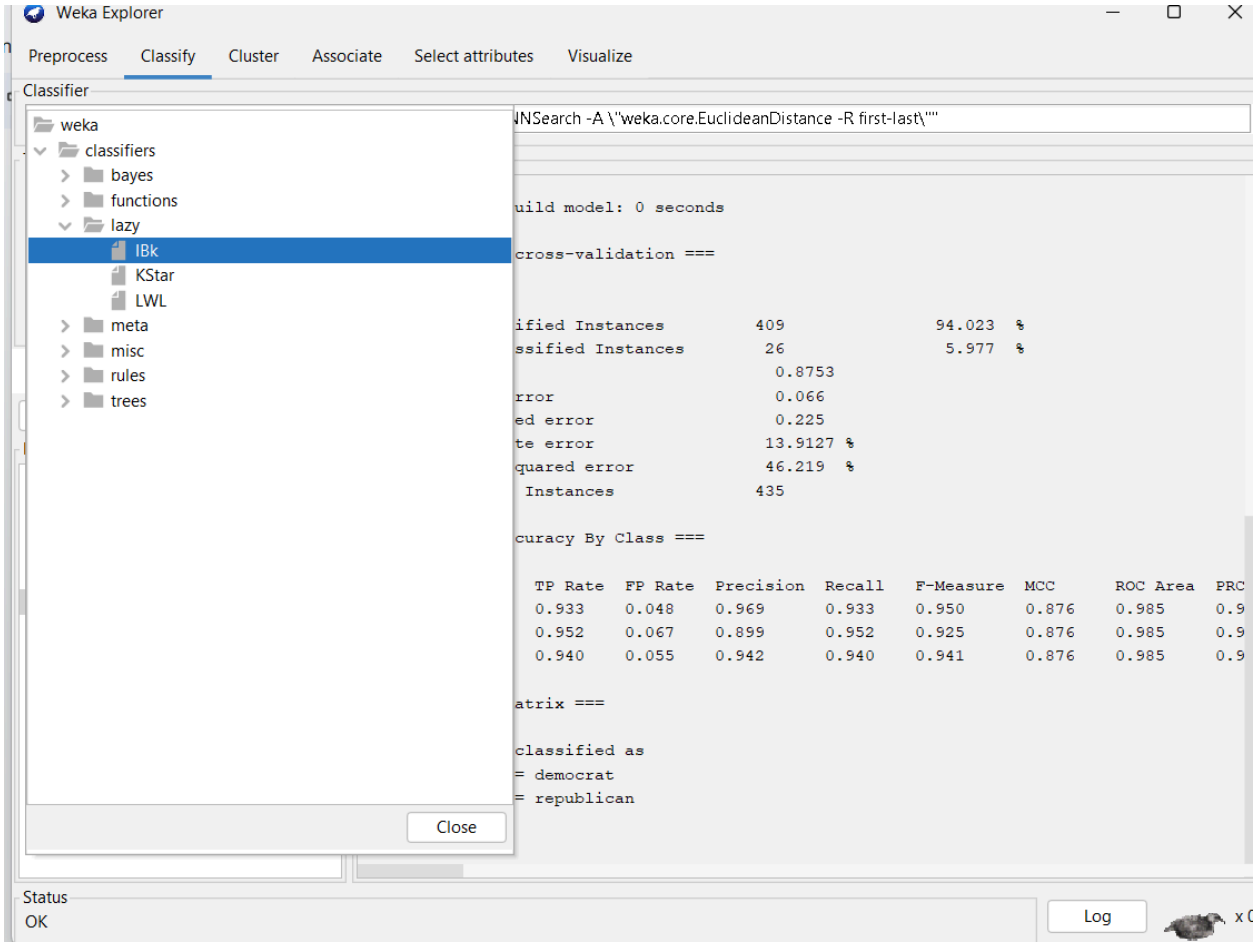


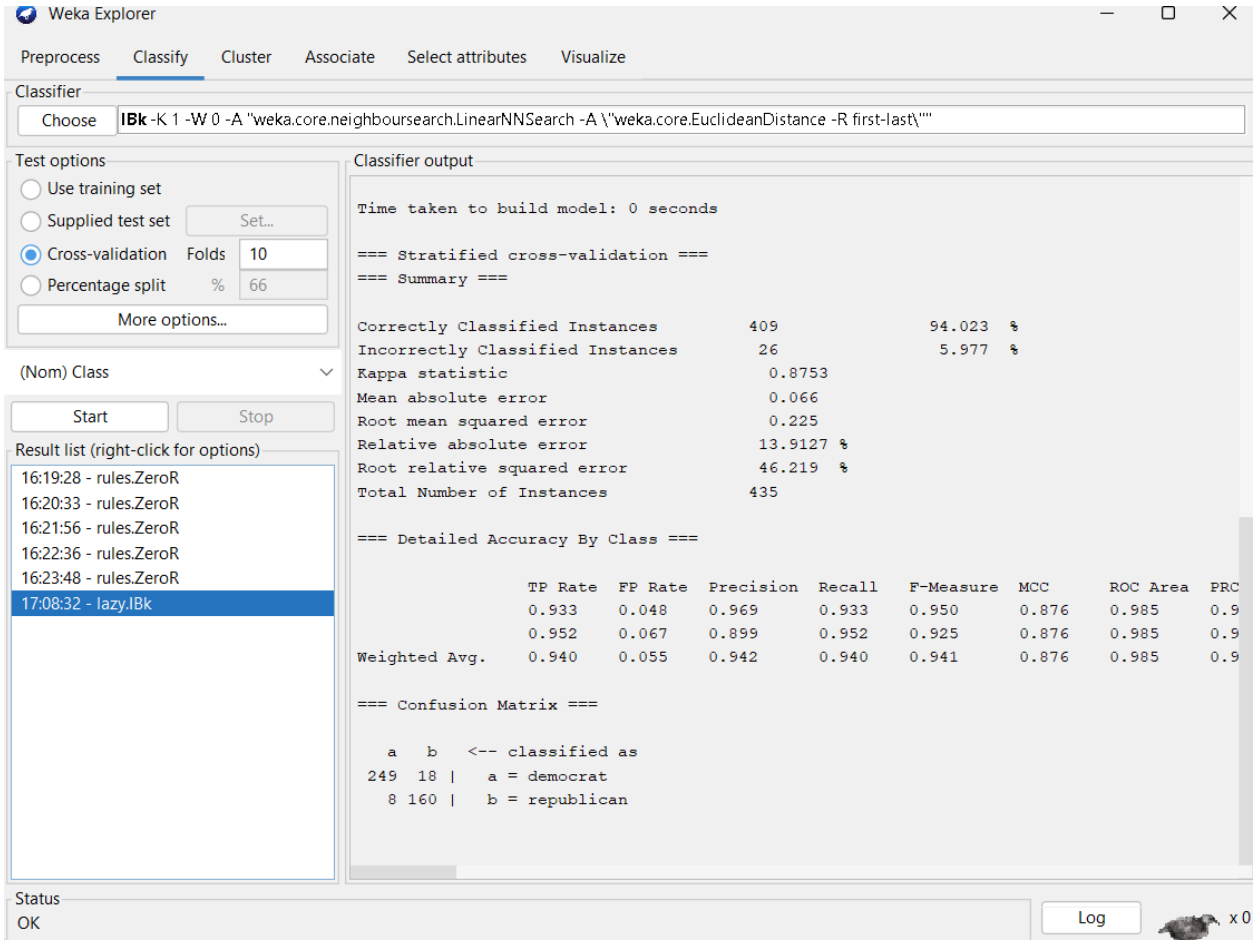
3. **Fact Constellation Schema:**



10. Implement classification of data using K-nearest neighbor approach in WEKA

1. Load Dataset:
    - Open WEKA.
    - Go to the "Explorer" tab.
    - Click "Open file" and select your dataset.
2. Choose KNN Algorithm:
    - In the "Classify" tab, select "IBk" from the list.
3. Set Options:
    - Click "Choose" and set the value of "K".
    - Adjust other options like distance metric if needed.
4. Evaluate Model:
    - Click "Start" to build the KNN model.
    - WEKA displays evaluation results (accuracy, precision, etc.).
5. Interpret Results:
    - Analyze evaluation results and visualize classification results.
6. Save Model (Optional):
    - Click "Save model" to save the trained model.





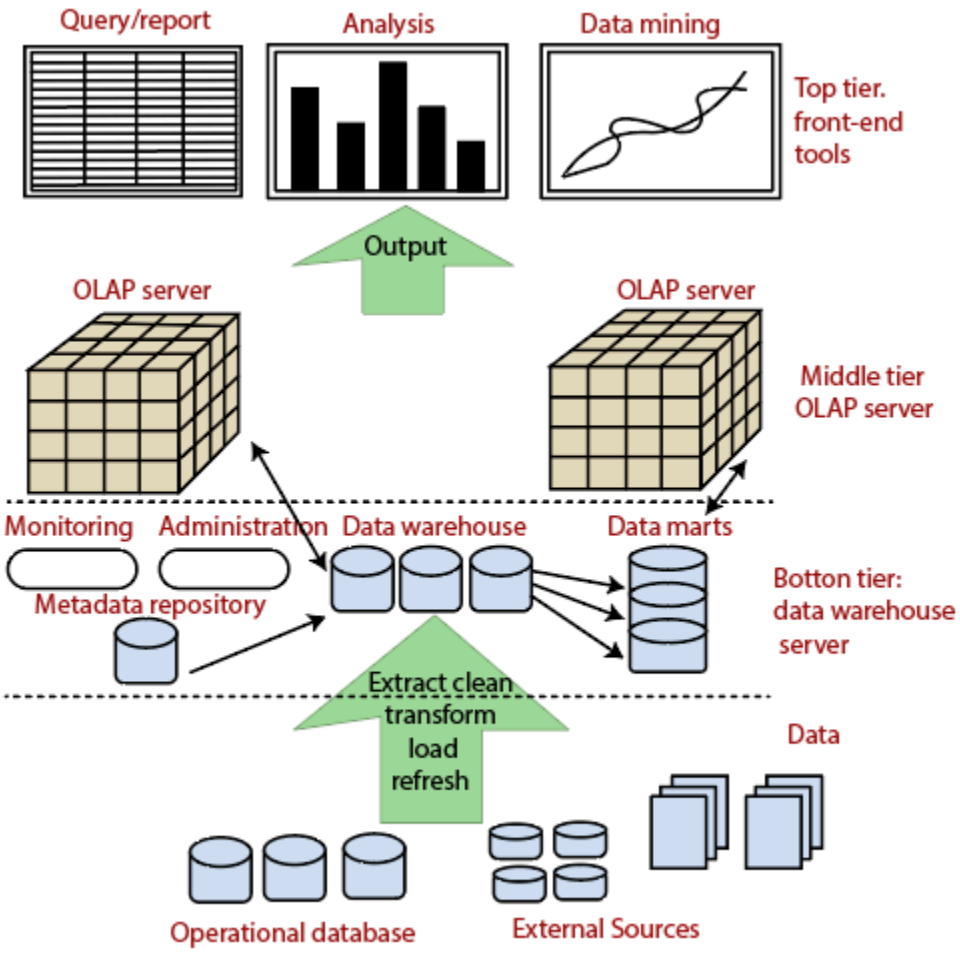
11. Design a three-tier data ware house architecture of banking.
And
12. Design a three-tier data ware house architecture of your own example.

Three-tier data warehouse architecture for banking:

1. **Data Sources Tier:**
    - Includes various data sources such as transaction systems, customer databases, and external data feeds.
    - Data is extracted, transformed, and loaded (ETL) into the data warehouse staging area.
2. **Data Warehouse Tier:**
    - Central repository for integrated and consolidated data from different sources.
    - Consists of the data warehouse database, where data is organized into dimensional or relational schemas.

- Supports reporting, analytics, and decision-making processes.

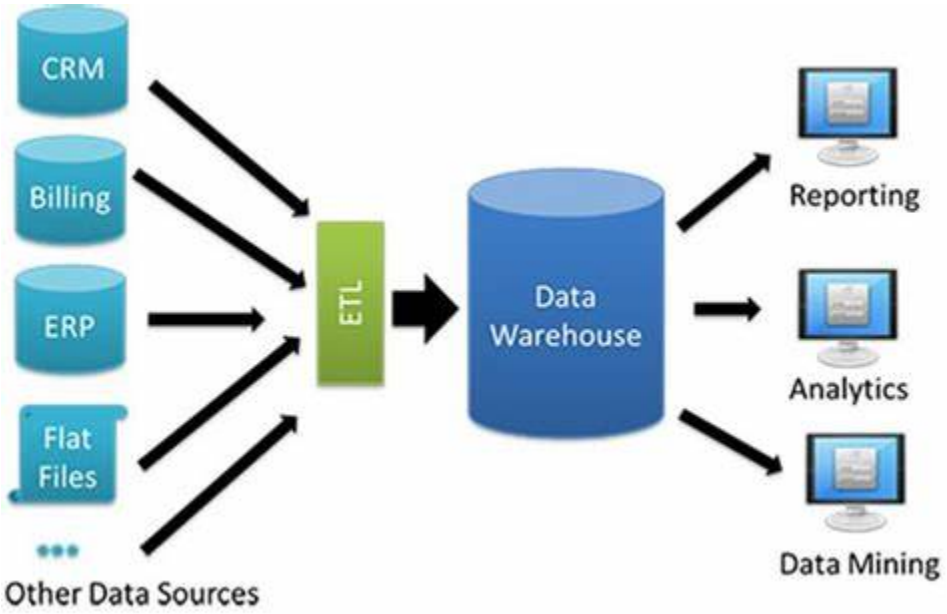
3. **Data Access Tier:**
    - Front-end tools and applications for accessing and analyzing data.
    - Includes dashboards, reporting tools, OLAP cubes, and data mining applications.
    - Users interact with data through intuitive interfaces to gain insights and make informed decisions.



Three-Tier Data Warehouse Architecture

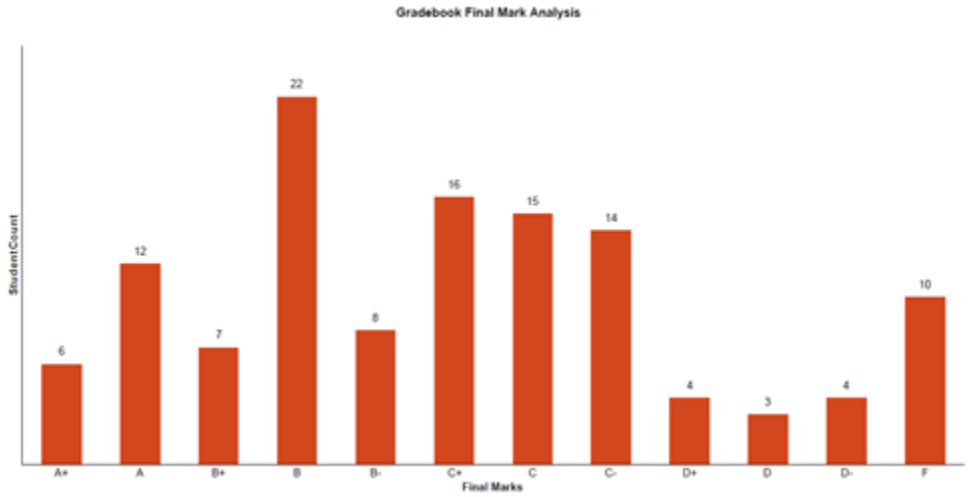
13. Build data warehouse system for storing and analyzing data about all loans issued by bank.

1. **Data Sources:**
    - Loan origination systems, customer databases, and external credit bureaus provide loan data.
    - ETL processes extract, transform, and load loan data into the data warehouse staging area.
2. **Data Warehouse:**
    - Central repository stores loan data in dimensional or relational schemas.
    - Includes a loan fact table with loan attributes (loan amount, interest rate, term) and related dimension tables (customer, branch, time).
    - Historical loan data is maintained for trend analysis and reporting.
3. **Data Access:**
    - BI tools and reporting dashboards provide access to loan data for analysis.
    - Users can track loan performance, analyze default rates, and identify patterns using OLAP cubes and data mining tools.
    - Ad-hoc queries and predefined reports facilitate decision-making and risk assessment.



14. Analyze the results of its students and the quality of its courses over time for a university.

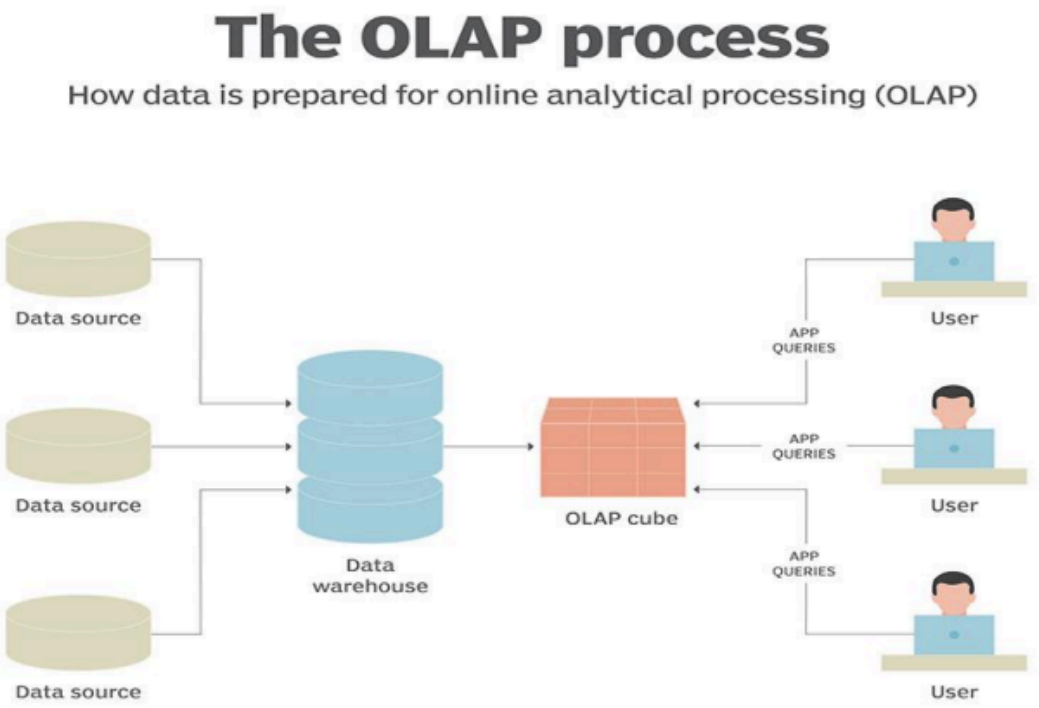
1. **Data Sources:**
    - Student information systems, course management systems, and survey data provide student and course information.
    - ETL processes integrate and load data into the data warehouse.
2. **Data Warehouse:**
    - Central repository stores student and course data in relational or dimensional schemas.
    - Includes a student fact table with attributes such as GPA, credits earned, and enrollment status.
    - Course dimension table contains details like course code, title, instructor, and enrollment statistics.
3. **Data Access:**
    - BI tools and reporting dashboards enable analysis of student performance and course quality trends over time.
    - Users can track graduation rates, dropout rates, course completion rates, and student satisfaction scores.
    - Insights help optimize curriculum, identify areas for improvement, and enhance teaching methodologies.



15. Perform a OLAP case study for sales analysis of a retail chain.

1. **Data Sources:**
    - Transactional systems capture sales data including product sales, customer information, and store details.
    - ETL processes extract, transform, and load data into the data warehouse.
2. **Data Warehouse:**
    - Central repository stores sales data in a star schema.
    - Fact table contains sales metrics such as revenue, quantity sold, and profit margin.
    - Dimension tables include product, customer, time, and store dimensions.
3. **OLAP Cube Design:**
    - Create a sales cube with dimensions such as product, customer, time, and store.
    - Measure sales metrics including revenue, quantity sold, and profit margin.
    - Define hierarchies for each dimension (e.g., product category, customer segment, date hierarchy).
4. **Analysis Queries:**
    - Analyze sales performance by product category, customer segment, and store location.
    - Compare sales trends over time, identify seasonality patterns, and evaluate sales growth.
    - Drill down into specific product lines, customer demographics, or store performance to uncover insights.
5. **Visualization and Reporting:**
    - Use BI tools to create interactive dashboards and reports.
    - Visualize sales trends, compare performance across dimensions, and identify outliers.
    - Share insights with stakeholders through scheduled reports and ad-hoc analysis.
6. **Decision-Making:**
    - Utilize insights to optimize inventory management, marketing strategies, and store operations.
    - Identify top-selling products, high-value customers, and underperforming stores for targeted interventions.

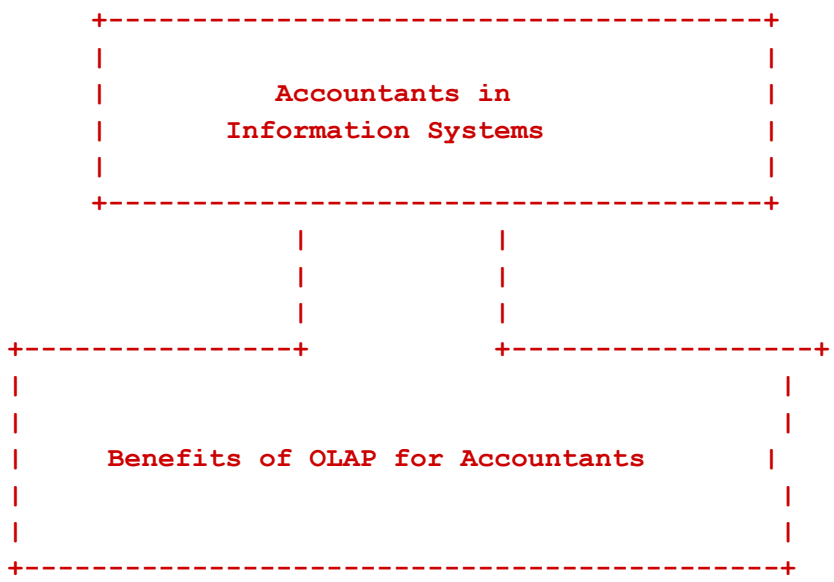
- Monitor key performance indicators (KPIs) and track progress towards sales targets.



16. Design and investigate role of accountants in information systems and the benefits of using OLAP to accountants in generating complex management reports

1. **Role:**
   - Accountants play a crucial role in designing, implementing, and maintaining information systems (IS) within organizations.
   - They ensure accurate financial data recording, compliance with regulatory requirements, and effective internal controls.
   - Accountants collaborate with IT professionals to develop accounting software, ERP systems, and reporting tools tailored to financial needs.
2. **Benefits of OLAP for Accountants:**

- OLAP (Online Analytical Processing) enables accountants to analyze large volumes of financial data quickly and efficiently.
- Accountants can generate complex management reports by drilling down into multidimensional data cubes to uncover insights.
- OLAP allows for flexible reporting, enabling accountants to create customized reports based on specific dimensions such as time, product, or customer.
- Accountants can perform variance analysis, trend analysis, and forecasting to support decision-making processes and strategic planning.
- OLAP empowers accountants to identify patterns, anomalies, and opportunities for cost reduction or revenue optimization.
- Overall, OLAP enhances the analytical capabilities of accountants, enabling them to provide valuable insights to management and stakeholders.

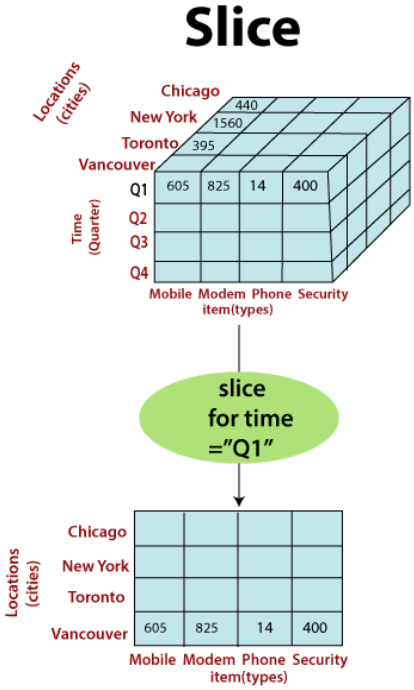


17. Perform Various OLAP operations such slice, dice, roll up, drill up and pivot.

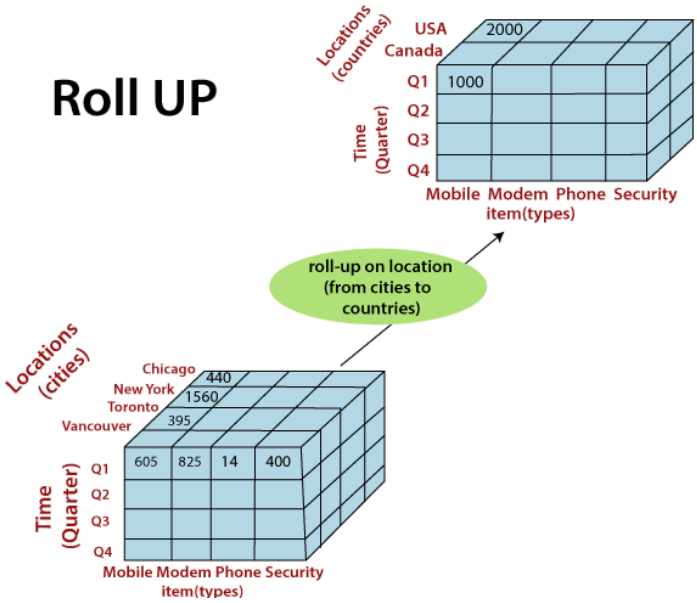
Various OLAP Operations:

1. **Slice:**

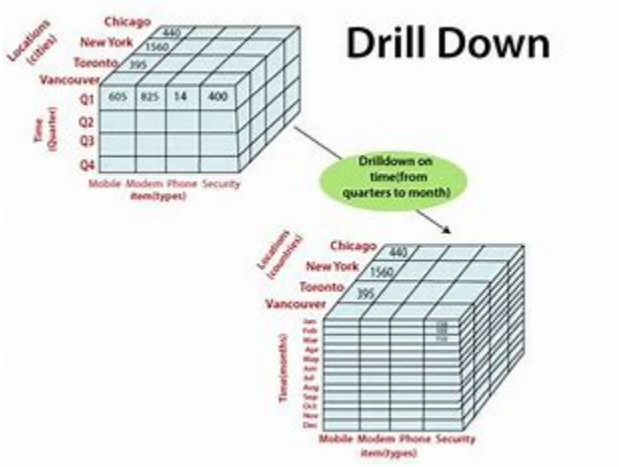
- Focus on a subset of data by selecting specific values for one or more dimensions.
- Example: Slicing sales data to only include sales from a particular region or time period.



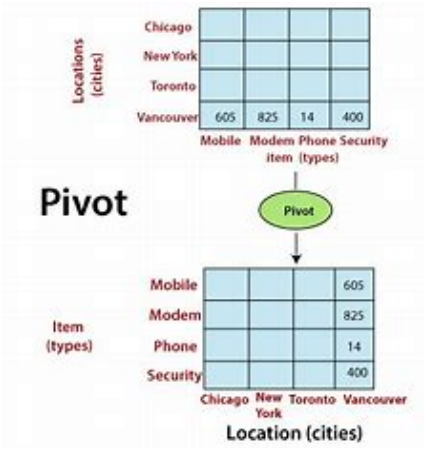
2. **Dice:**
    - Zoom in on a specific portion of data by selecting multiple values for multiple dimensions.
    - Example: Dicing sales data to include sales from a specific region and product category during a certain time frame.
3. **Roll Up:**
    - Aggregate data across one or more dimensions to a higher level of abstraction.
    - Example: Rolling up sales data from daily to monthly or yearly totals.



4. **Drill Up:**
    - Navigate from detailed data to summarized data by moving up the hierarchy of one or more dimensions.
    - Example: Drilling up from daily sales data to view sales by week, month, or year.



5. **Pivot:**
    - Rotate or reorient the data to view it from a different perspective.
    - Example: Pivoting sales data to switch the rows and columns, allowing for easier comparison or analysis.



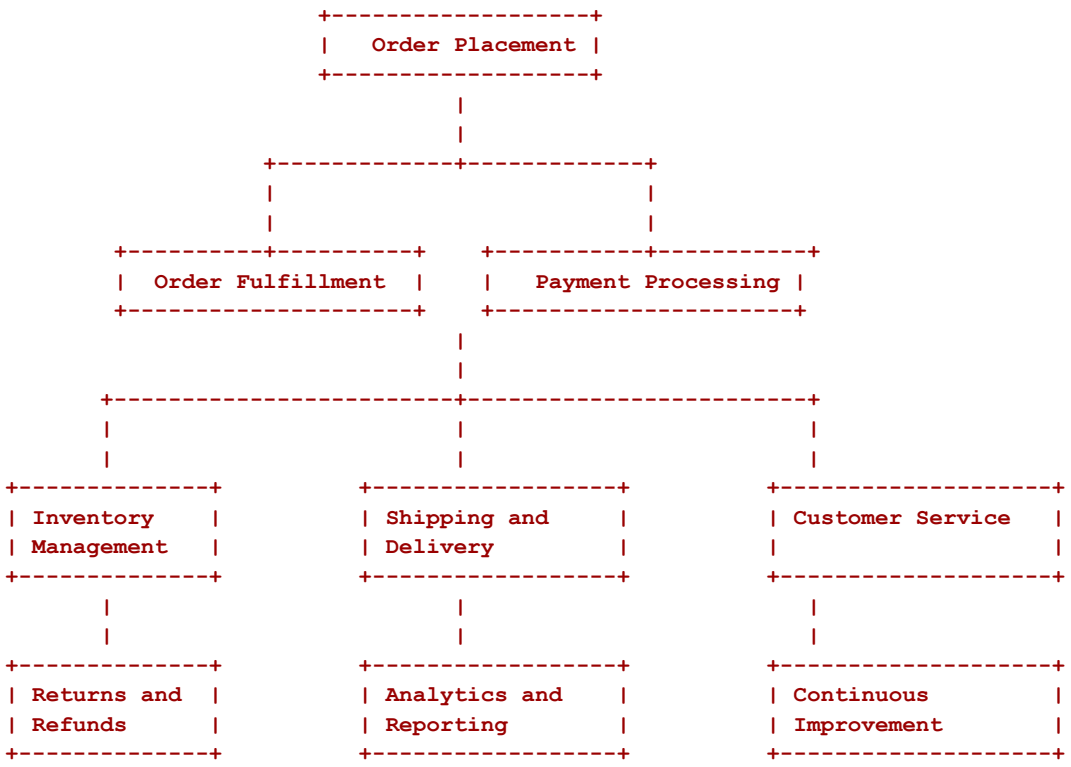
18. Perform a case study on online retail order processing

Online Retail Order Processing Case Study:

1. **Order Placement:**
    - Customers browse the online store, add items to their cart, and proceed to checkout.
    - They provide shipping and payment information and submit the order.
2. **Order Fulfillment:**
    - The order details are transmitted to the warehouse for picking, packing, and shipping.
    - Warehouse staff pick items from inventory, pack them securely, and generate shipping labels.
3. **Inventory Management:**
    - Inventory levels are continuously monitored to ensure availability of products.
    - Replenishment orders are placed with suppliers to restock popular items.
4. **Payment Processing:**
    - Payment gateway processes credit card transactions securely and verifies payment authorization.
    - Once payment is confirmed, the order status is updated to "payment received".
5. **Shipping and Delivery:**
    - Orders are dispatched from the warehouse and handed over to shipping carriers.
    - Customers receive tracking information to monitor the status of their shipments.
6. **Customer Service:**

- Customer service representatives handle inquiries, resolve issues, and provide assistance with orders.
- They communicate with customers via email, phone, or chat to ensure satisfaction.

7. **Returns and Refunds:**
   - Customers may initiate returns or request refunds for damaged, incorrect, or unwanted items.
   - Return requests are processed, and refunds are issued upon receipt of returned merchandise.
8. **Analytics and Reporting:**
   - Data on sales, orders, inventory levels, and customer behavior is analyzed to identify trends and patterns.
   - Reports on key performance indicators (KPIs) such as order fulfillment time, customer satisfaction, and sales revenue are generated to guide business decisions.
9. **Continuous Improvement:**
   - Feedback from customers and performance metrics are used to identify areas for improvement in the order processing workflow.
   - Process optimizations and technology enhancements are implemented to streamline operations and enhance the customer experience.



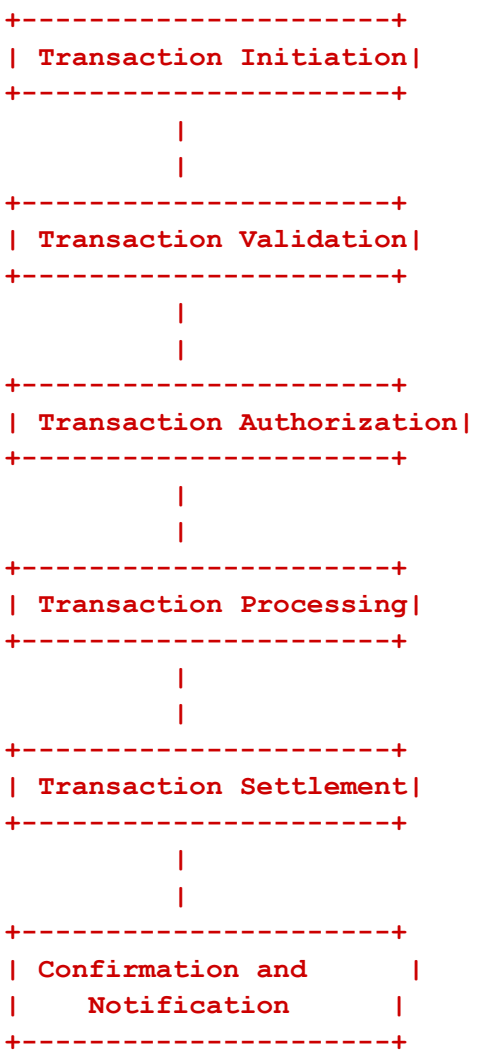
19. Performa banking transaction processing system.

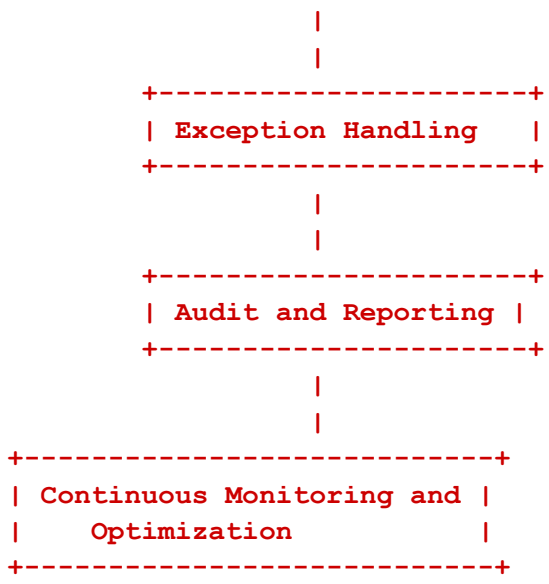
Banking Transaction Processing System:

1. **Transaction Initiation:**
    - Customers initiate transactions through various channels such as ATM, online banking, mobile banking, or in-person at branches.
    - They provide transaction details such as account numbers, transaction type, and amount.
2. **Transaction Validation:**
    - The system validates transaction requests against customer account information, transaction limits, and security protocols.
    - Valid transactions proceed to the next stage, while invalid transactions are rejected.
3. **Transaction Authorization**:
    - Authorized personnel review and approve transactions based on predefined rules and authorization levels.
    - Multi-factor authentication may be required for high-value or sensitive transactions.
4. **Transaction Processing:**
    - Approved transactions are processed, updating account balances and transaction records in real-time.
    - Funds are transferred between accounts, and transaction details are recorded for audit and reporting purposes.
5. **Transaction Settlement:**
    - Transactions are settled through the appropriate clearing and settlement systems, such as Automated Clearing House (ACH) or Real-Time Gross Settlement (RTGS).
    - Settlement ensures the transfer of funds between financial institutions and finalizes transaction processing.
6. **Confirmation and Notification:**
    - Customers receive confirmation of successful transactions through email, SMS, or online banking notifications.
    - Confirmation includes transaction details, updated account balances, and reference numbers for tracking.
7. **Exception Handling:**

- The system identifies and resolves exceptions such as insufficient funds, transaction errors, or fraud alerts.
- Exception handling processes ensure the integrity and security of banking transactions.

8. **Audit and Reporting:**
    - Transaction data is recorded in transaction logs and databases for audit trail and reporting purposes.
    - Reports on transaction volumes, trends, and exceptions are generated to monitor system performance and compliance.
9. **Continuous Monitoring and Optimization:**
    - The system is continuously monitored for performance, security, and compliance with regulatory requirements.
    - Ongoing optimization efforts ensure efficient transaction processing and enhance the customer experience.

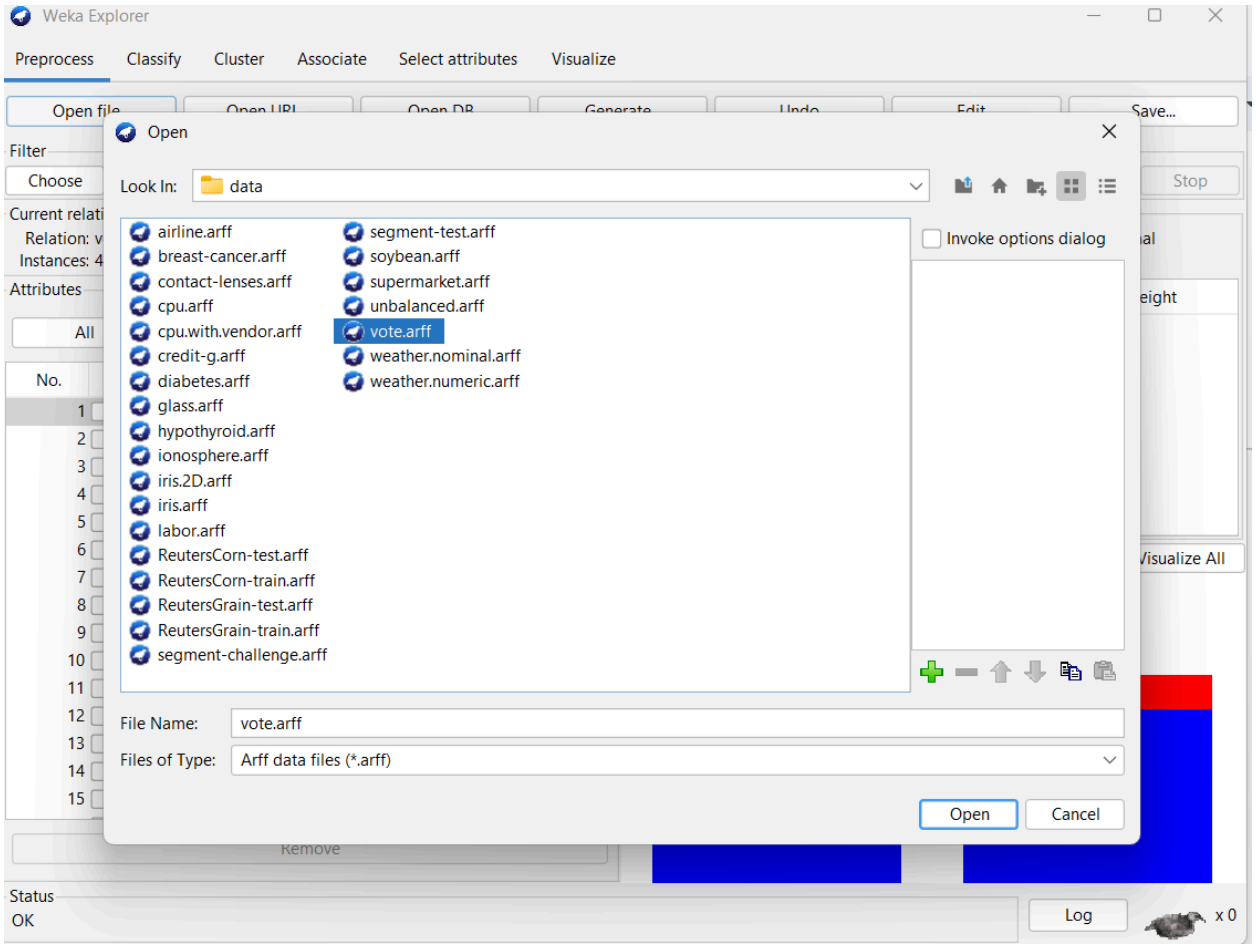


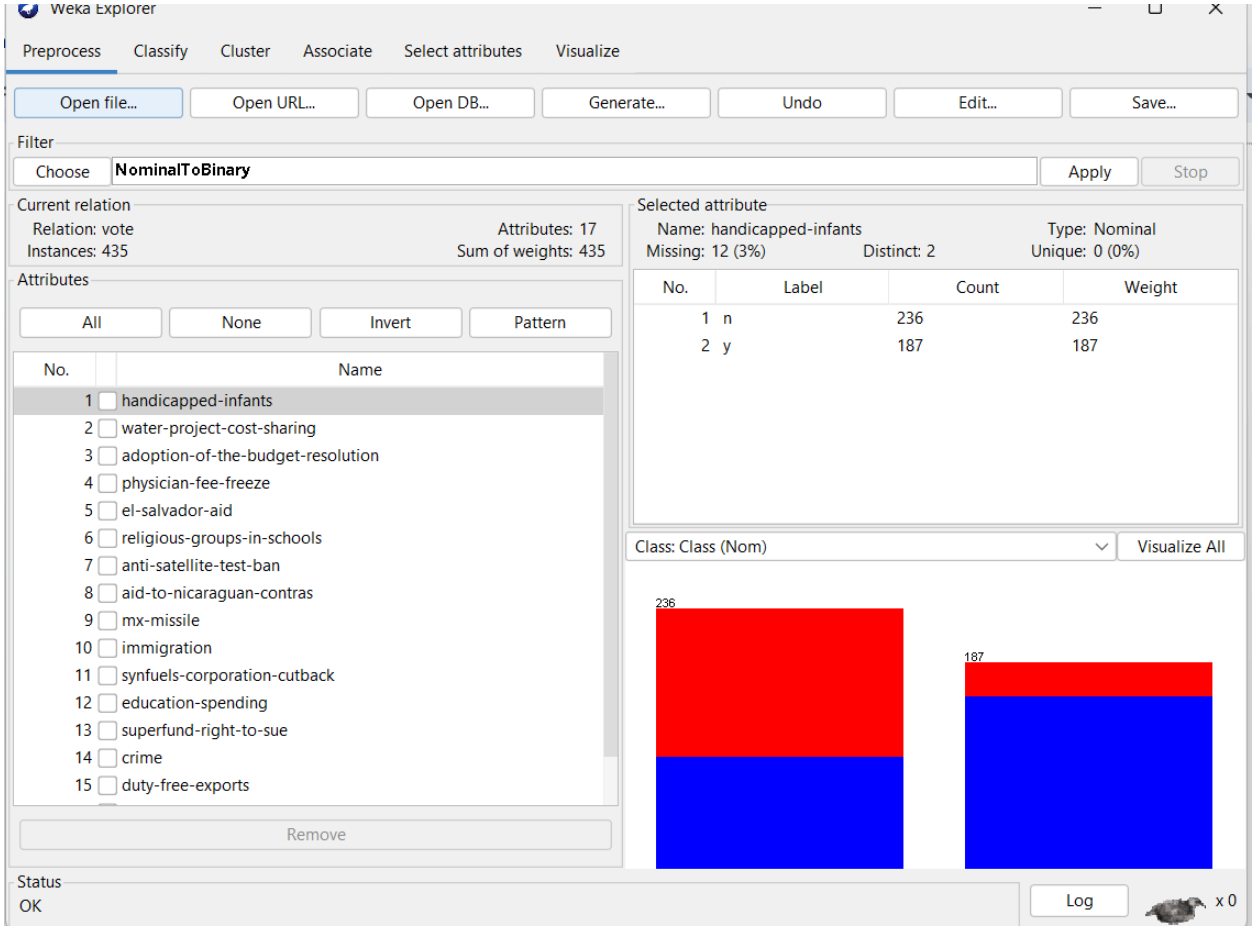


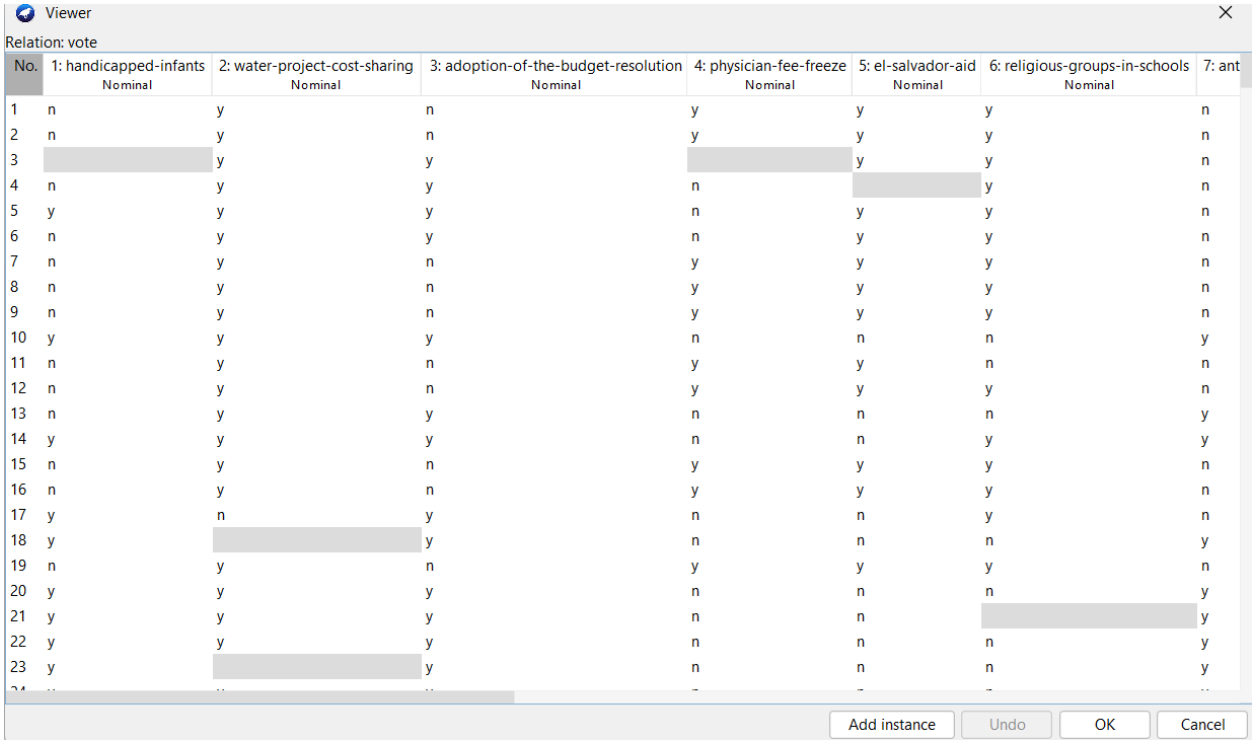
20. Implement warehouse testing.

1. **Load Data:**
   - Open WEKA and load the dataset extracted from the data warehouse.
2. **Explore Data:**
   - Use WEKA's Explorer tab to explore the dataset, examining attributes, instances, and summary statistics.
3. **Preprocess Data:**
   - Preprocess the data as needed, including handling missing values, transforming attributes, and filtering noise.
4. **Select Test Method:**
   - Choose the appropriate testing method based on your objectives. Common methods include:
     - Classification: Predicting a class label or category for each instance.
     - Regression: Predicting a numerical value for each instance.
     - Clustering: Grouping similar instances into clusters based on their attributes.
     - Association rule mining: Discovering patterns and relationships between attributes.
5. **Choose Evaluation Metrics:**
   - Select evaluation metrics to assess the performance of the test method. Common metrics include accuracy, precision, recall, F-measure, and ROC curves.

6. **Split Data:**
    - Split the dataset into training and testing subsets to evaluate the model's performance. Cross-validation techniques can also be employed for more robust evaluation.




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
NominalToBinary
Apply
Stop
Current relation
Relation: vote
Instances: 435
Attributes: 17
Sum of weights: 435
Selected attribute
Name: handicapped-infants
Missing: 12 (3%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No. Label Count Weight
1 n 236 236
2 y 187 187
Attributes
All
None
Invert
Pattern
No. Name
1 handicapped-infants
2 water-project-cost-sharing
3 adoption-of-the-budget-resolution
4 physician-fee-freeze
5 el-salvador-aid
6 religious-groups-in-schools
7 anti-satellite-test-ban
8 aid-to-nicaraguan-contras
9 mx-missile
10 immigration
11 synfuels-corporation-cutback
12 education-spending
13 superfund-right-to-sue
14 crime
15 duty-free-exports
Remove
Class: Class (Nom)
Visualize All
236
187
Status
OK
Log
x 0


Viewer
Relation: vote
No.
1: handicapped-infants
Nominal
2: water-project-cost-sharing
Nominal
3: adoption-of-the-budget-resolution
Nominal
4: physician-fee-freeze
Nominal
5: el-salvador-aid
Nominal
6: religious-groups-in-schools
Nominal
7: ant
1 n y n y y y n
2 n y n y y y n
3 y y y y n
4 n y y n y n
5 y y y n y y n
6 n y y n y y n
7 n y n y y y n
8 n y n y y y n
9 n y n y y y n
10 y y y n n n y
11 n y n y y n n
12 n y n y y y n
13 n y y n n n y
14 y y y n n y y
15 n y n y y y n
16 n y n y y y n
17 y n y n n y n
18 y y n n n y
19 n y n y y y n
20 y y y n n n y
21 y y y n n y
22 y y y n n n y
23 y y n n n y
Add instance
Undo
OK
Cancel