🎯 Retail Intelligence Platform - Final Submission Guide

Team BNZ - Snowflake x Accenture Hackathon

✅ Complete Implementation Checklist

Phase 1: Environment Setup □

- □ Snowflake trial account created (US West - Oregon)
- □ Database `RETAIL_INTELLIGENCE_DB` created
- □ Schemas created: RAW, STAGING, ANALYTICS, MART
- □ Warehouse `RETAIL_WH` created and configured
- □ ACCOUNTADMIN role used for permissions

Phase 2: Data Ingestion □

- □ Downloaded Abt-Buy dataset from https://dbs.uni-leipzig.de/files/datasets/Abt-Buy.zip
- □ Created raw tables: ABT_PRODUCTS, BUY_PRODUCTS, PRODUCT_MAPPING_TRUTH
- □ Created file format and stage
- □ Loaded all CSV files successfully
- □ Verified data counts match expected values
- □ Data quality checks passed

Phase 3: Data Transformation □

- □ DBT project setup (or staging views created)
- □ STG_ABT_PRODUCTS created with cleaned data
- □ STG_BUY_PRODUCTS created with cleaned data
- □ Price conversions working correctly
- □ Text normalization applied (UPPER, TRIM)
- □ Null handling implemented

Phase 4: Feature Engineering with Snowpark □

- □ Snowpark Python environment configured
- □ Custom UDFs created (Jaccard similarity, Levenshtein distance)
- □ Product pair features table created
- □ Features include: name similarity, description similarity, price differences
- □ Token overlap calculated
- □ Feature validation completed

Phase 5: AI/ML with Cortex □

- □ **PRODUCT_EMBEDDINGS** table created using EMBED_TEXT_768
- □ Embeddings generated for all products (ABT + BUY)
- □ **PRODUCT_MATCHES_AI** table created with cosine similarity
- □ Match confidence levels assigned (HIGH, MEDIUM, LOW)
- □ Match ranking implemented (top N per product)
- □ **PRODUCT_CATEGORIES** table created using COMPLETE LLM
- □ AI categorization applied to all products
- □ Semantic similarity threshold tuned (≥ 0.55)

Phase 6: Price Intelligence □

- □ **PRICE_INTELLIGENCE** view created
- □ Price differences calculated (absolute & percentage)
- □ Price positions identified (cheaper/parity)
- □ Competitiveness scores computed
- □ Category-level price analysis available
- □ Price opportunity identification working

Phase 7: Model Validation □

- □ **MATCH_VALIDATION** view created
- □ Ground truth comparison implemented
- □ Precision metrics calculated by confidence level
- □ **MODEL_PERFORMANCE** view created
- □ **OVERALL_PERFORMANCE** metrics available
- □ Accuracy > 70% for HIGH confidence matches
- □ Coverage metrics tracked

Phase 8: Cortex Analyst □

- □ Semantic model YAML file created
- □ All tables properly defined in YAML
- □ Dimensions and measures configured
- □ Relationships defined
- □ Sample questions included
- □ YAML uploaded to CORTEX_ANALYST_STAGE
- □ Cortex Analyst tested with natural language queries
- □ At least 5 test questions validated

Phase 9: Snowflake Intelligence Dashboards □

- □ Dashboard created in Snowsight
- □ Key metrics tiles added (match statistics, prices)
- □ Visualizations created (charts, graphs)

- □ Match confidence distribution chart
- □ Price position breakdown chart
- □ Category analysis charts
- □ Model performance metrics displayed
- □ Dashboard filters configured
- □ Screenshots captured

Phase 10: Streamlit Application □

- □ Streamlit app created in Snowflake
- □ Connection to Snowflake configured
- □ **Product Matching** tab implemented
- □ **Price Intelligence** tab implemented
- □ **Category Analysis** tab implemented
- □ **Search Products** tab implemented
- □ **Model Performance** tab implemented
- □ **AI Assistant** tab with Cortex Analyst
- □ Interactive filters working
- □ Data export functionality added
- □ Error handling implemented
- □ App tested end-to-end

📊 Snowflake Features Used - Verification

Required Features ✅

Feature	Used	Evidence
Snowpark	✅	Custom Python UDFs for similarity calculations
DBT Projects	✅	Staging and analytics models for data transformation
Cortex AI - Embeddings	✅	EMBED_TEXT_768 for semantic product vectors
Cortex AI - Complete	✅	LLM for product categorization
Cortex Agents	✅	Cortex Analyst for natural language queries
AI/SQL	✅	VECTOR_COSINE_SIMILARITY for product matching
Snowflake Intelligence	✅	Interactive dashboards in Snowsight
Streamlit	✅	Full-featured web application

Additional Features Used 🌟

- **Vector Search**: Semantic similarity using embeddings
- **Dynamic Tables**: (Optional) For real-time updates
- **Stages**: Internal stage for data loading
- **Views**: Analytics and mart layer views
- **UDFs**: Custom text processing functions
- **Time Travel**: Data versioning capability

📁 Deliverables Checklist

1. Code Repository ☐

```
retail-intelligence-platform/
├── README.md (comprehensive with setup instructions)
├── architecture_diagram.png
├── sql/
│   ├── 01_setup.sql
│   ├── 02_data_loading.sql
│   ├── 03_transformations.sql
│   ├── 04_ml_features.sql
│   ├── 05_cortex_ai.sql
│   └── 06_analytics.sql
├── snowpark/
│   ├── feature_engineering.py
│   ├── requirements.txt
│   └── test_features.py
├── dbt/ (if used)
│   ├── dbt_project.yml
│   └── models/
├── cortex/
│   ├── retail_semantic_model.yaml
│   └── test_analyst.py
├── streamlit/
│   ├── app.py
│   └── requirements.txt
├── dashboards/
│   ├── dashboard_queries.sql
│   └── screenshots/
├── tests/
│   ├── data_quality.sql
│   └── validation.sql
└── docs/
    ├── setup_guide.md
    ├── user_guide.md
    └── technical_documentation.md
```

2. Documentation □

☐ **README.md** with:

- Problem statement
- Solution overview
- Architecture diagram
- Setup instructions
- Features used
- Results and metrics
- Team information

☐ **Setup Guide** with:

- Step-by-step installation
- Prerequisites
- Configuration details
- Troubleshooting section

☐ **Technical Documentation** with:

- Data model diagrams
- SQL query explanations
- ML model details
- API documentation

☐ **User Guide** with:

- Dashboard navigation
- Streamlit app usage
- Cortex Analyst queries
- Feature descriptions

3. Presentation Materials □

☐ **PowerPoint/PDF Slides** (10-15 slides):

- Title slide with team name
- Problem statement

- Solution architecture
- Technical implementation
- Snowflake features showcase
- Demo screenshots
- Results and metrics
- Business value
- Future enhancements
- Q&A slide

□ **Demo Video** (3-5 minutes):

- Introduction (30 sec)
- Problem & solution overview (1 min)
- Live demo of key features (2-3 min)
- Results and impact (30 sec)
- Conclusion (30 sec)
- Professional quality audio/video
- Screen recordings of actual application

4. Screenshots & Visuals □

□ Architecture diagram
□ Data flow diagram
□ Snowsight dashboard screenshots (at least 3)
□ Streamlit app screenshots (all tabs)
□ Cortex Analyst query examples
□ SQL query examples with results
□ Model performance charts
□ Price intelligence visualizations

5. Test Results □

□ Data quality test results
□ Model performance metrics:

- Overall accuracy: ___%
- Precision by confidence level
- Recall metrics
- Coverage statistics

□ Performance benchmarks:

- Query response times

- Embedding generation time
- Dashboard load time

□ Sample outputs:

- Top 10 product matches
- Price analysis results
- Category distribution
- AI-generated insights

🎯 Key Metrics to Highlight

Quantitative Results

1. **Data Scale**
    - Total products processed: ___
    - Total matches generated: ___
    - High confidence matches: ___
2. **Model Performance**
    - Overall precision: ___%
    - HIGH confidence precision: ___%
    - MEDIUM confidence precision: ___%
    - Recall rate: ___%
3. **Business Insights**
    - Average price difference: ___%
    - Products with >10% savings: ___
    - Categories analyzed: ___
    - Price opportunities identified: ___
4. **Technical Performance**
    - Average query time: ___ ms
    - Embedding generation time: ___ sec
    - Dashboard load time: ___ sec
    - Concurrent users supported: ___

Qualitative Results

1. **Innovation**
   - Multi-strategy matching approach
   - AI-powered categorization
   - Natural language query interface
   - Real-time price intelligence
2. **User Experience**
   - Intuitive dashboard design
   - Interactive filtering
   - Natural language queries
   - Export capabilities
3. **Business Value**
   - Automated product matching
   - Competitive pricing insights
   - Market intelligence
   - Decision support

🚀 Submission Guidelines

Before Submitting

1. **Final Testing**
   - Run all SQL scripts from scratch
   - Test Streamlit app end-to-end
   - Verify all dashboards load correctly
   - Test Cortex Analyst with 10+ queries
   - Check all exports and downloads
2. **Code Quality**
   - Add comments to complex queries
   - Include error handling
   - Use consistent formatting
   - Remove hardcoded credentials

- Add logging where appropriate

3. **Documentation Review**
    - Check for typos and grammar
    - Verify all links work
    - Ensure screenshots are clear
    - Update version numbers
    - Add contact information

4. **Performance Optimization**
    - Add appropriate indexes
    - Optimize slow queries
    - Cache frequently used results
    - Review warehouse sizing

Submission Checklist

□ All code pushed to GitHub/GitLab
□ README.md is complete and clear
□ Documentation is comprehensive
□ Presentation slides finalized
□ Demo video recorded and uploaded
□ Screenshots captured and organized
□ Test results documented
□ Requirements.txt/dependencies listed
□ No sensitive data or credentials in code
□ License file included (if required)

What to Submit

1. **GitHub Repository URL** with:
    - Public access enabled
    - Clear README
    - All code and documentation

2. **Presentation Deck** (PDF format)

3. **Demo Video** (MP4, hosted on YouTube/Vimeo or included in repo)

4. **Snowflake Account Details** (for judges to test):
    - Account URL
    - Read-only username/password

- Database and schema names
- Dashboard URLs
- Streamlit app URL

5. **Submission Form** with:
    - Team name: BNZ
    - Use case: Retail Intelligence Platform
    - Features used: (list all)
    - GitHub URL
    - Video URL
    - Any special instructions

🌟 Tips for Standing Out

Technical Excellence

1. **Advanced Features**
    - Implement multiple matching strategies
    - Use ensemble methods for categorization
    - Add confidence calibration
    - Implement A/B testing framework
2. **Performance Optimization**
    - Optimize vector similarity calculations
    - Use materialized views strategically
    - Implement caching layers
    - Add query profiling
3. **Code Quality**
    - Write clean, documented code
    - Use design patterns
    - Implement comprehensive testing
    - Add CI/CD pipeline

User Experience

1. **Dashboard Design**
   - Professional visualizations
   - Intuitive navigation
   - Responsive layout
   - Helpful tooltips
2. **Streamlit App**
   - Modern UI/UX
   - Fast response times
   - Error messages that guide users
   - Export functionality
3. **Documentation**
   - Clear setup instructions
   - Visual diagrams
   - Use case examples
   - Troubleshooting guide

Business Impact

1. **Quantify Value**
   - Cost savings identified
   - Time saved vs manual matching
   - Accuracy improvement
   - Scale achieved
2. **Real-World Application**
   - Explain use cases
   - Show scalability
   - Discuss deployment strategy
   - Address data privacy
3. **Future Roadmap**
   - Enhancement ideas
   - Additional features
   - Integration possibilities

- Scaling strategy

📞 Support Resources

Snowflake Documentation

- Cortex AI: https://docs.snowflake.com/en/user-guide/snowflake-cortex
- Snowpark: https://docs.snowflake.com/en/developer-guide/snowpark/python/index
- Streamlit: https://docs.streamlit.io/
- DBT: https://docs.getdbt.com/

Community Support

- Snowflake Community: https://community.snowflake.com/
- Stack Overflow: Tag [snowflake-cloud-data-platform]
- GitHub Discussions

Hackathon Specific

- Review hackathon rules and judging criteria
- Check submission deadline
- Verify required features are used
- Confirm submission format

🏆 Final Checklist Before Submission

☐ All features from requirements are implemented
☐ Code runs without errors
☐ Documentation is complete
☐ Presentation is polished
☐ Demo video is ready
☐ Test results are documented
☐ Repository is public
☐ All team members reviewed
☐ Submission form filled out
☐ Deadline confirmed

□ Backup copy saved

Good Luck! 🎉

You've built an impressive AI-powered Retail Intelligence Platform using Snowflake's cutting-edge features. Make sure to:

1. **Test everything** one final time
2. **Document thoroughly** - judges love clear docs
3. **Show business value** - not just technical prowess
4. **Practice your demo** - first impressions matter
5. **Submit early** - don't wait until the last minute

Remember: Your solution solves a real business problem using advanced AI and modern data engineering practices. Be confident and highlight the innovation!

Questions or Issues?

If you encounter any problems:

1. Check the Snowflake documentation
2. Review error messages carefully
3. Test queries in isolation
4. Use Snowflake's query profiler
5. Check community forums

Team BNZ - You've got this! 🚀