**📊 Pair Plot Analysis**

This analysis is based on the pair plot generated from the following selected features:

* numofproducts
* gender\_label
* isactivemember
* balancerange
* estimatedsalaryrange
* creditscorerange
* tenurerange
* age\_skewed

**🧐 Key Observations:**

**1. numofproducts vs. other features**

* The distribution is highly discrete — very few unique values (like 1, 2, 3, 4).
* Customers with more products (3 or 4) seem rare.
* No strong visible relationship between numofproducts and balance, salary, or credit score.

**Recommendation**:  
✅ Keep for ML models (especially tree-based models).  
✳️ Not ideal for linear models without encoding.

**2. gender\_label vs. other features**

* Gender is a binary variable (0 or 1).
* No clear patterns with balance, credit score, or age in the scatter plots.

**Recommendation**:  
🔹 Optional — very weak correlation, might not improve models much.

**3. isactivemember vs. other features**

* Binary variable.
* Customers who are *active members* show slight distribution differences in balancerange and estimatedsalaryrange.
* Active members may have slightly better credit scores.

**Recommendation**:  
✅ Keep — activity status could be important in predicting churn or customer behavior.

**4. balancerange vs. other features**

* Customers mostly have **either a very low balance or a very high balance** (bimodal distribution).
* No strong linear relationship with salary or credit score.
* Might be informative together with geography.

**Recommendation**:  
✅ Keep for modeling — balance is a strong financial indicator.

**5. estimatedsalaryrange vs. other features**

* Very wide spread.
* No clear relationship with credit score, age, or tenure.

**Recommendation**:  
🔹 Optional — keep for tree models, may not help linear models much.

**6. creditscorerange vs. other features**

* Credit score values are concentrated at specific bands.
* Slight trend: customers with higher scores might have slightly lower products and balances.

**Recommendation**:  
✅ Keep — credit score remains critical for risk-based predictions.

**7. tenurerange vs. other features**

* Very little visible pattern with other numerical features.
* Slight hint that longer tenure doesn't strongly correlate with higher balance or salary.

**Recommendation**:  
✅ Keep — tenure could help models capture retention/churn trends.

**8. age\_skewed vs. other features**

* Clear patterns: older customers have different distributions compared to younger customers.
* Slight negative trend with balance and credit score.

**Recommendation**:  
✅ Very important — Age is consistently predictive for customer behavior.

**🔥 Overall Recommendations:**

| **Column** | **Recommendation** | **Notes** |
| --- | --- | --- |
| numofproducts | ✅ Keep | Good for tree-based models. |
| gender\_label | 🔹 Optional | Might not add much value. |
| isactivemember | ✅ Keep | Important for behavior modeling. |
| balancerange | ✅ Keep | Financial indicator. |
| estimatedsalaryrange | 🔹 Optional | Keep for tree models. |
| creditscorerange | ✅ Keep | Key for customer evaluation. |
| tenurerange | ✅ Keep | Helpful for churn analysis. |
| age\_skewed | ✅ Keep | Critical feature. |

**🎯 Final Tip:**

* For **linear models**, you might drop gender\_label and estimatedsalaryrange.
* For **tree models (Random Forest, XGBoost, LightGBM)** — keep all features.

