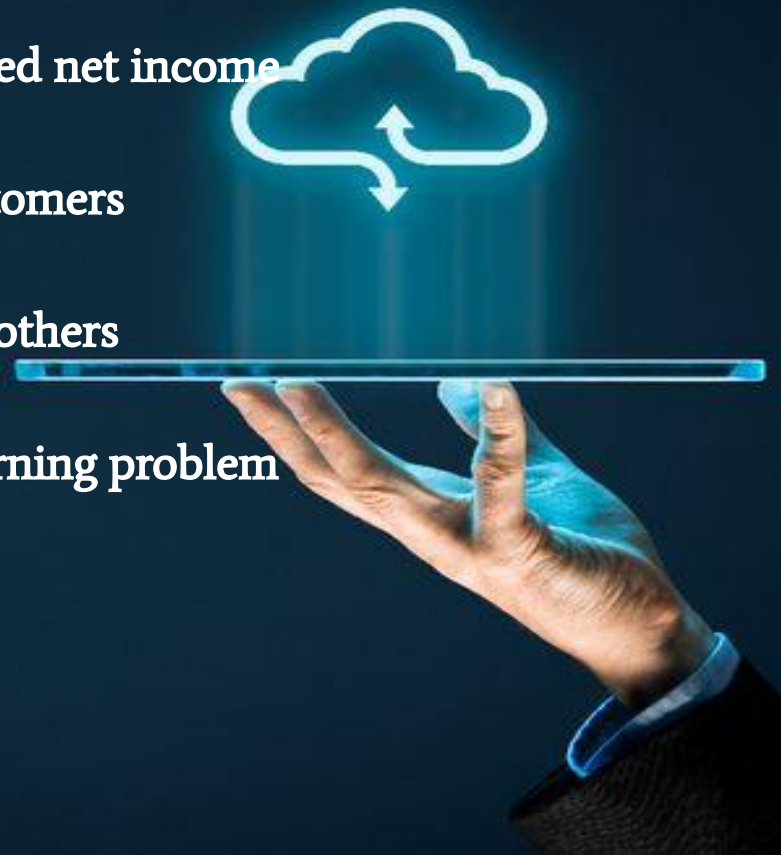

95.55 %
20.00 %
33.55 %
15.00 %
945.0 -1.15 93.36 % 95,096
20.55 +1.95 4.50 % 85
15.95 +0.2 33.55 % 588,955

Sprint Focus

- Build a machine learning pipeline
- Predict customer income
- Use transaction history and customer data

Problem We Are Solving

- We want to predict a customer's declared net income
- Income is already known for some customers
- We use this data to predict income for others
- This makes it a regression machine learning problem



Data Overview

- 14 months of customer transactions
- Customer demographic information
- Employment-related data
- One customer can have many transactions
- Income exists once per customer

Data Preparation

- Cleaned missing and incorrect values
- Grouped transactions by customer
- Converted transaction activity into summaries

Feature Engineering

- Total amount spent
- Average transaction value
- Number of transactions
- Highest and lowest transaction values

Modeling & Evaluation

- Trained a baseline regression model
- Used RMSE to measure accuracy
- RMSE shows how close predictions are to actual income