SUPERANNUATION FUND DATA ANALYSIS

Welcome to our comprehensive analysis of the Australian Superannuation Fund member dataset. This presentation provides senior management and data analysts with key insights into our member data, investment patterns, and data quality issues that require attention.

Our analysis covers data quality metrics, missing value patterns, outlier detection, standardisation needs, and exploratory findings that reveal important relationships between member demographics, contributions, and superannuation balances.

SR **by Stewart Robertson**

INITIAL DATA ASSESSMENT

DATASET OVERVIEW

500 rows across 20 columns containing member personal and financial information including demographics, contribution rates, and investment options

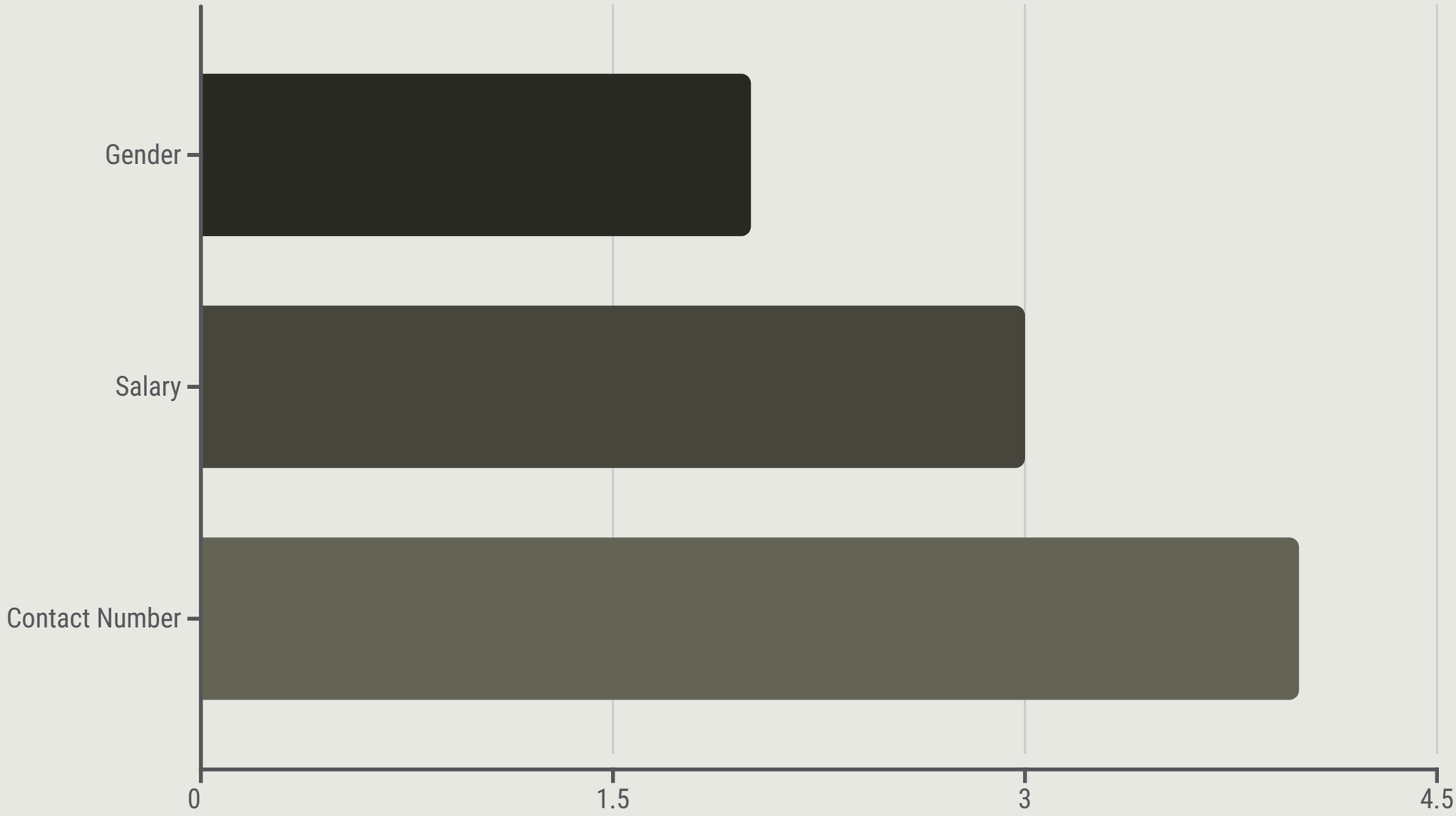
DATA QUALITY ISSUES

Missing values in gender (rows 6, 7), salary (rows 5, 6, 16), and contact numbers (rows 5, 8, 17, 19)

INTEGRITY CHECK

No duplicates found based on member_id, but potential inconsistencies in gender representation and date formatting

MISSING VALUE ANALYSIS



Our analysis identified several missing values that require attention. For gender, we recommend imputation using the mode or based on employment status patterns. Salary gaps can be addressed by using median values for respective employment categories.

Contact number missing values can be handled with "N/A" placeholders. These strategies will ensure data completeness while maintaining analytical integrity for downstream processing and reporting.

OUTLIER DETECTION AND HANDLING

SALARY OUTLIERS

Z-Score: 2 outliers

IQR: 1 outlier

Recommendation: Cap at 95th percentile or apply log transformation to reduce skewness

CONTRIBUTION RATE OUTLIERS

Employer: 1 IQR outlier

Employee: 1 IQR outlier

Recommendation: Further investigation needed as these may be valid policy-driven values

SUPER BALANCE OUTLIERS

IQR: 1 outlier

Recommendation: Cap extreme values or transform to reduce skewness while preserving data integrity

DATA STANDARDISATION REQUIREMENTS

Column Name	Current Format	Required Format
date_of_birth	MM/DD/YYYY	YYYY-MM-DD
gender	Various representations	Standardised categories
contact_number	Various formats	XXX-XXX-XXXX
insurance_coverage	String (true/false)	Boolean (True/False)
salary, contribution rates	Float	Float with two decimal places

Standardisation is crucial for accurate analysis. We recommend converting all dates to YYYY-MM-DD format, creating a consistent mapping for gender representation, and ensuring numeric values maintain consistent decimal precision.

EXPLORATORY DATA ANALYSIS INSIGHTS

$253K

AVERAGE SALARY

With median of $250K and standard deviation of $100K

0.12

MEAN EMPLOYER RATE

Average employer contribution rate across members

0.05

MEAN EMPLOYEE RATE

Average employee contribution rate across members

$5M

MEAN SUPER BALANCE

With median of $4.5M and standard deviation of $2M

Our analysis reveals significant variation in member salaries and super balances. The high standard deviations indicate the presence of outliers that may skew analyses. The data shows a diverse membership base with varying contribution patterns that warrant further investigation.

KEY CORRELATION FINDINGS

SALARY & SUPER BALANCE

Strong positive correlation (0.430) between salary and super balance, indicating higher earners have significantly larger retirement savings

OVERALL IMPACT

Multi-faceted influences on superannuation balances requiring further exploration of demographic factors

EMPLOYER CONTRIBUTIONS

Moderate correlation (0.430) between salary and employer contribution rate, suggesting higher earners receive better employer contributions

EMPLOYEE CONTRIBUTIONS

Weak correlation (0.176) between salary and employee contribution rate, with minimal relationship to super balance (-0.045)

RECOMMENDATIONS & NEXT STEPS

DATA CLEANING IMPLEMENTATION

Implement proposed strategies for handling missing values, standardising formats, and addressing outliers

DEEPER ANALYSIS

Investigate relationships between employment status, age demographics, and super balance growth

ENHANCED VISUALISATION

Develop comprehensive visualisations to effectively communicate insights on member demographics and investment patterns

STRATEGIC INSIGHTS DEVELOPMENT

Explore impact of age on contribution rates and retirement readiness to inform future fund policies