

臺灣證券交易所建置大數據 平台之經驗分享

資訊服務部 黃俊豪

2016年2月26日

大綱

- 1 目的
- 2 平台建置
- 3 業務導入
- 4 資料匯入
- 5 進度追蹤
- 6 成果分享
- 7 結論

目的

需求

討論

開發

測試

執行



需求排程處理



待處理後取得結果



調整時重複作業



目的(續)

建置平台

- 提供自主性分析平台
- 減少作業流程
- 避免需求與結果不一致

自行分析

- 自行分析需求
- 快速得到分析結果
- 可重複調整需求

降低負荷

- 減少處理需求負擔
- 專心處理專案

平台建置

平台評選

BIG
DATA

訂定
題目

邀集
廠商

提供
資料

廠商
簡報

投票
評選

議價
採購

軟硬體
建置



平台擴充

平台建置(續)

評選題目

上市公司

- 103年第2季財報流動比率大於80%，應收帳款週轉天數小於150天，現金流量比率大於0
- 102年7月至103年6月營收較101年7月至102年6月營收成長

1

權證交易

- 請分別計算各家證券商於103. Q2及103. Q3之報價情形季分數及等級
- 報價情形季分數及等級計算規則

2

股票交易

- 以102年7、8月【揭示檔】及102年8月底【發股數、上市股數、收盤價】按(1)依市值分大、中、小(2)依週轉率(7、8月成交值合計/8月底市值)分高、中、低(3)權證、非權證分析

3

平台建置(續)

評選題目

ETF

- 分析成交檔中，證券商對ETF下單之交易情形

4

市場監理

- 參數畫面
- 個股買(賣)前N名投資人
- 群組期間交易行為分析
 - 群組日內交易行為分析
 - 群組日內交易氣泡分析
- 產製完整分析報表檔(Word)



5

平台建置(續)



評選流程

簡報順序抽籤決定

題目說明及問答

說明後不再接受電話詢問

廠商2天作業時間

每家廠商簡報及展示40分鐘，15分鐘答詢



平台建置(續)

投票評選

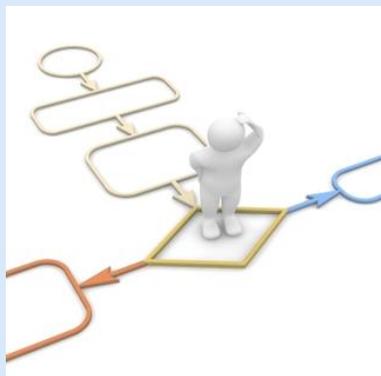
廠商介紹

逐題評分：正確性、操作流暢性、
畫面設計

總平均需達合格分數70分

總分相同情形，由主席裁決

9個部門組成委員會



平台建置(續)

竭誠為您服務

「導入視覺化統計分析工具」廠商評選表(總分法)

日期：103年11月21日

評選項目	內容	配分	參與評選廠商					
			廠商1	廠商2	廠商3	廠商4	廠商5	廠商6
簡報	廠商介紹	10						
題目一展示	正確性	15						
	操作流暢性							
	畫面設計							
題目二展示	正確性	15						
	操作流暢性							
	畫面設計							
題目三展示	正確性	15						
	操作流暢性							
	畫面設計							
題目四展示	正確性	15						
	操作流暢性							
	畫面設計							
題目五展示	正確性	15						
	操作流暢性							
	畫面設計							
答詢	相關問題討論	15						
評分合計		100						

評選單位：

評選委員簽名：

- 註：1. 評分後若投標廠商之總平均得分未達合格分數70分，不列入議價廠商。
 2. 加總計算各廠商之分數，由高至低為優先議價順序，若達合格廠商有總分相同情形，由主席裁決順序。
 3. 本表統計完成經評選委員及主席確認後，於當日下午17:30公布。

平台建置(續)

教育訓練

各部門一般人員

- 一般課程
- 6場次，300人次



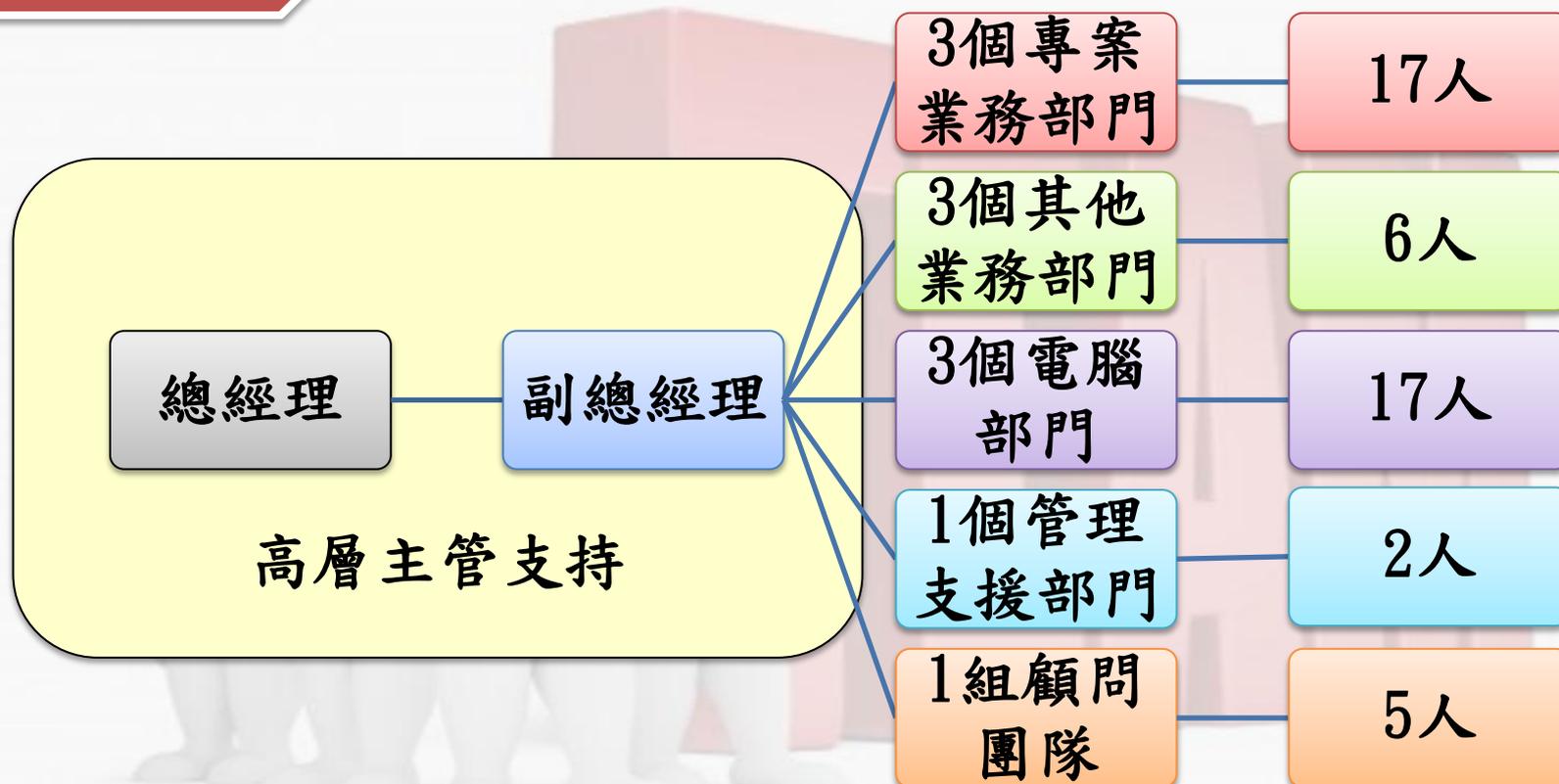
各部門種子人員

- 進階專業課程
- 平台管理課程



業務導入

專案組織



業務導入(續)

分析題目

1

• 投資人交易**股票**行為分析

68個子題

2

• 投資人交易認購(售)**權證**行為分析

61個子題

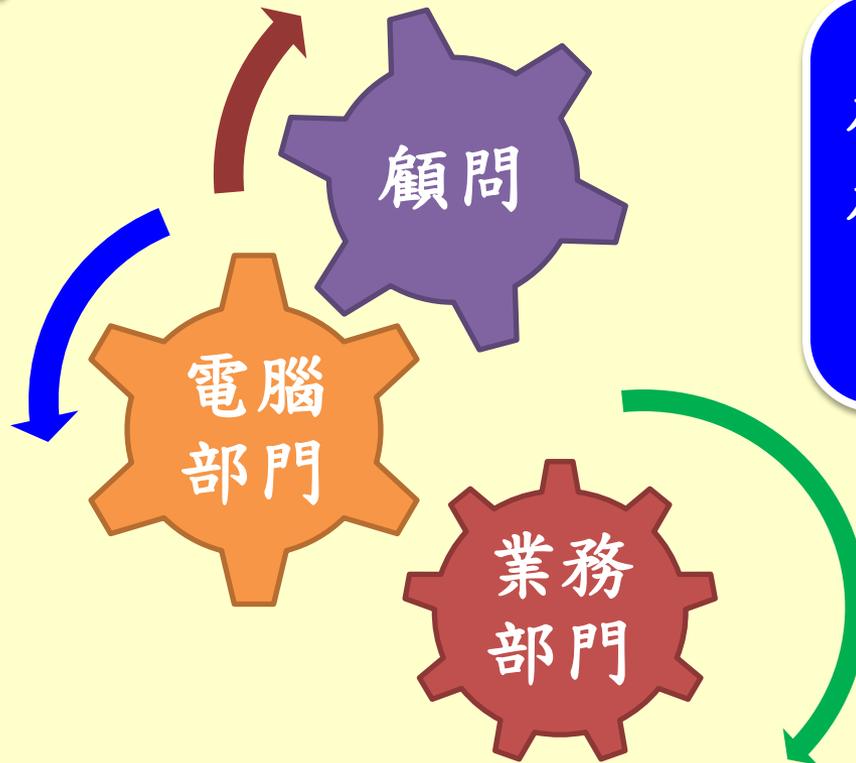
3

• 投資人交易**ETF**行為分析

62個子題

業務導入(續)

過程



確認分析資料
確定分析子題
資料彙整及合併

業務導入(續)

3個專案業務部門



自行分析及調整

報告撰寫

成果簡報

觀摩學習

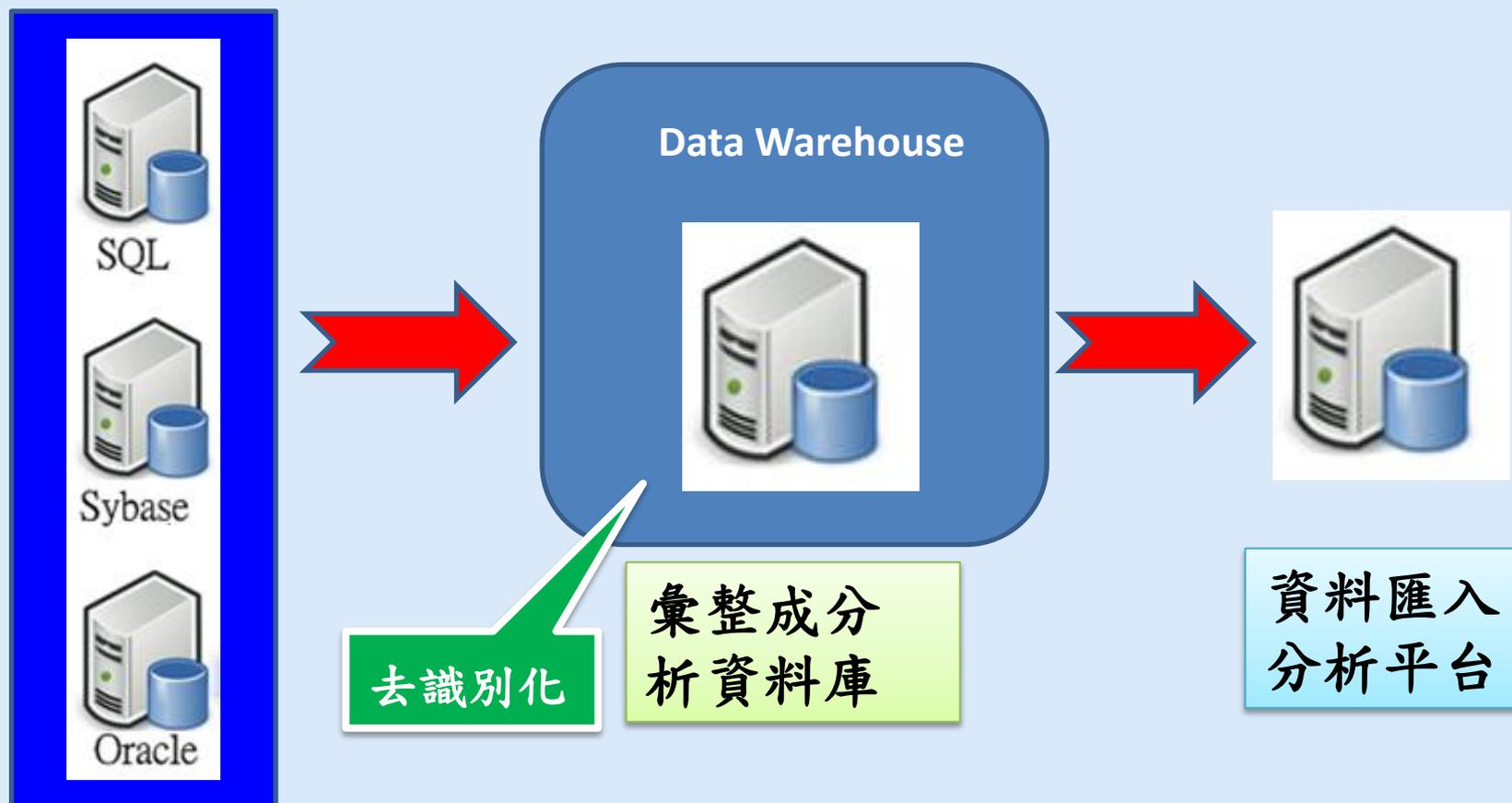
3個其他業務部門

資料匯入

資料擷取

資料轉換

資料匯入



去識別化

彙整成
分析資料庫

資料匯入
分析平台

資料匯入(續)

分析資料



期間

- 2010年1月至2015年6月
- 持續匯入至2016年2月

內容

- 開戶、證券、成交、信用交易、當沖、借券，共分14檔
- 月檔、季檔、3檔(股票、權證、ETF)

量

- 1.18T(in memory)
- 信用交易檔394G、成交檔363G

資料匯入(續)

資料整理

異質資料庫

- 連線函式庫不同
- 舊版本，外界可用資源較少
- 需自行找尋特殊參數或方式處理

資料架構

- 主檔及異動檔架構，資料轉換成每日資料
- 調整成統一架構並還原歷史軌跡

內容轉換清理

- 特殊字元訂定統一的資料格式
- 不斷調整、測試，發揮設計上創意



進度追蹤

視覺化分析平台建置



專案自2014年10月至2015年12月，共計15個月

專案完成

進度追蹤(續)

專案會議

進度會議

每周1次，
總經理
主持

2015年3
至11月，
共20次

主管機關

每季1次

2015年1
至12月，
共4次

業務討論

每周
1~2次

2015年1
至10月，
共28次

電腦討論

不定期

2015年1
月至12月，
共23次

成果分享

- 2015年11月12日發函提供各證券商資料
- 共計10項

股票

- 權值股及非權值股比重(季)
- 各地區性別及成交值比重(季)
- 信用交易／當沖占整體市場成交值比重(季)
- 信用交易權值股非權值股比重(季)
- 投資人日交易金額集中度檔(日)

權證

- 月投資人數分類檔(月)
 - 該月有交易權證-新增者
 - 該月有交易權證-持續投資者
 - 該月有交易權證-其他
 - 退場者

ETF

- 各類型ETF(台股ETF/國外ETF/槓反ETF/期貨ETF)資產規模月增減變化(月)
- 自然人對各類型ETF月成交金額(月)
- 自然人交易ETF之年齡分布(月)
- 自然人交易ETF之性別分布(月)

成果分享(續)

- 2016年1月置於本公司網站(www.twse.com.tw)
- [產品與服務] → [出版品] → [其他出品版] → [大數據統計資料(股票、權證、ETF)]



The screenshot shows the TWSE website interface. The navigation path is highlighted with red circles:

- Home → Products and Services (產品與服務)
- Products and Services → Publications (出版品)
- Publications → Other Publications (其他出品版)
- Other Publications → Big Data Statistical Information (大數據統計資料)

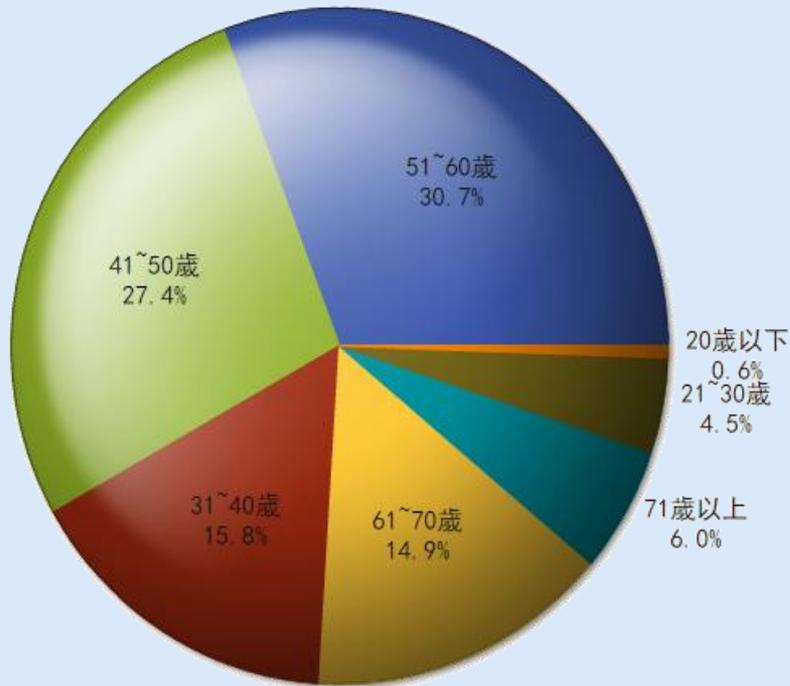
The 'Other Publications' table is shown below:

名稱	出刊日期	文件下載
證券資料彙編	月刊	15日
上市證券彙編	月刊	10日
有價證券借貸	年刊	六月
證券編碼	年刊	民國100年10月
證券服務	年刊	五月
查詢個人資料服務	年刊	三月
資訊服務	年刊	二月
電腦資訊交換平台	年刊	三月
市場發展新發展	年刊	三月
網路資訊商店	年刊	三月
出版品	年刊	三月
證券服務	年刊	三月
國際證券市場發展動態	年刊	三月
研究報告	年刊	三月
其他出版品	年刊	三月
投資人教育	年刊	三月
問答集	年刊	三月
證券統計彙編	年刊	二月
大數據統計資料(股票、權證、ETF)	特刊	2016年1月

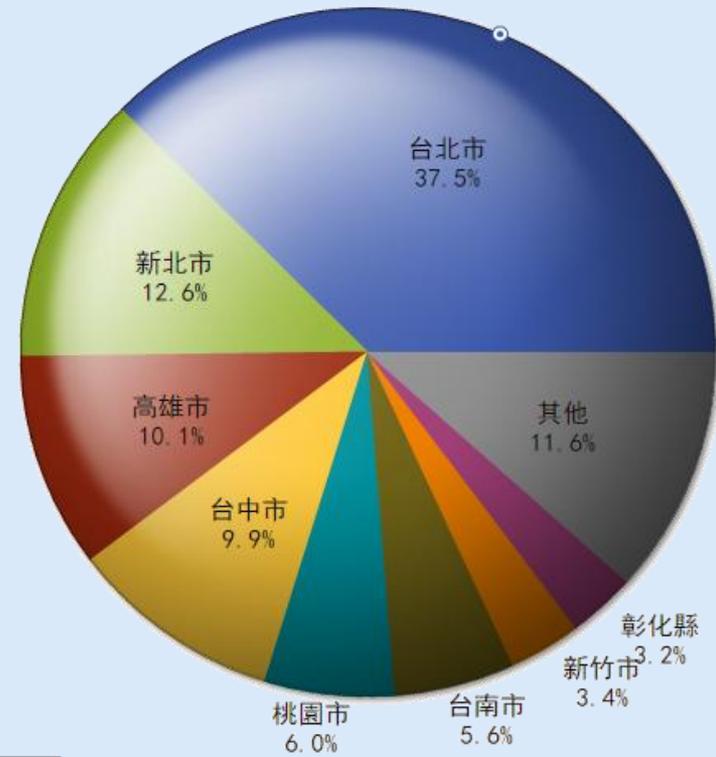
成果分享(續)

股票交易 ①

自然人各年齡級距成交額



自然人地區別成交金額

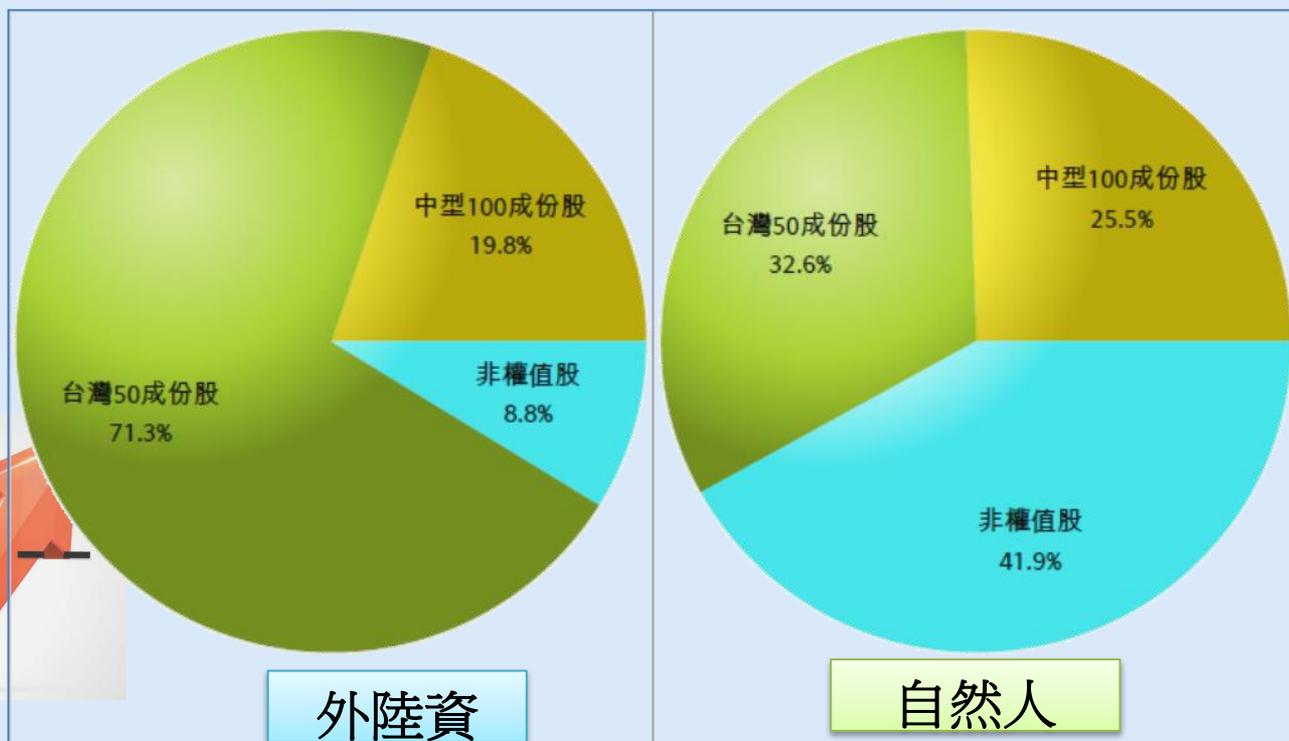


年齡級距

■ 20歲以下
 ■ 21~30歲
 ■ 31~40歲
 ■ 41~50歲
 ■ 51~60歲
 ■ 61~70歲
 ■ 71歲以上

成果分享(續)

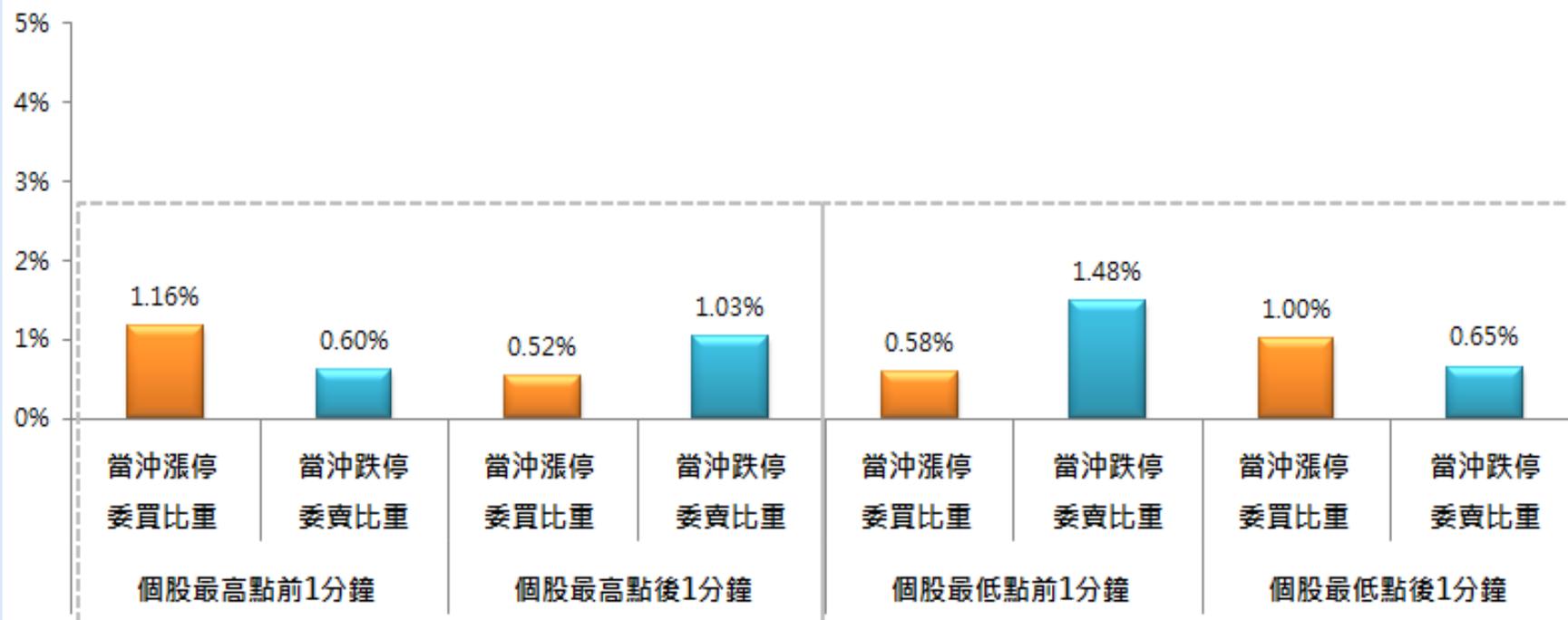
股票交易 ②



成果分享(續)

股票交易 ③

2015/6/1-9/30當沖前20大證券漲跌停委託



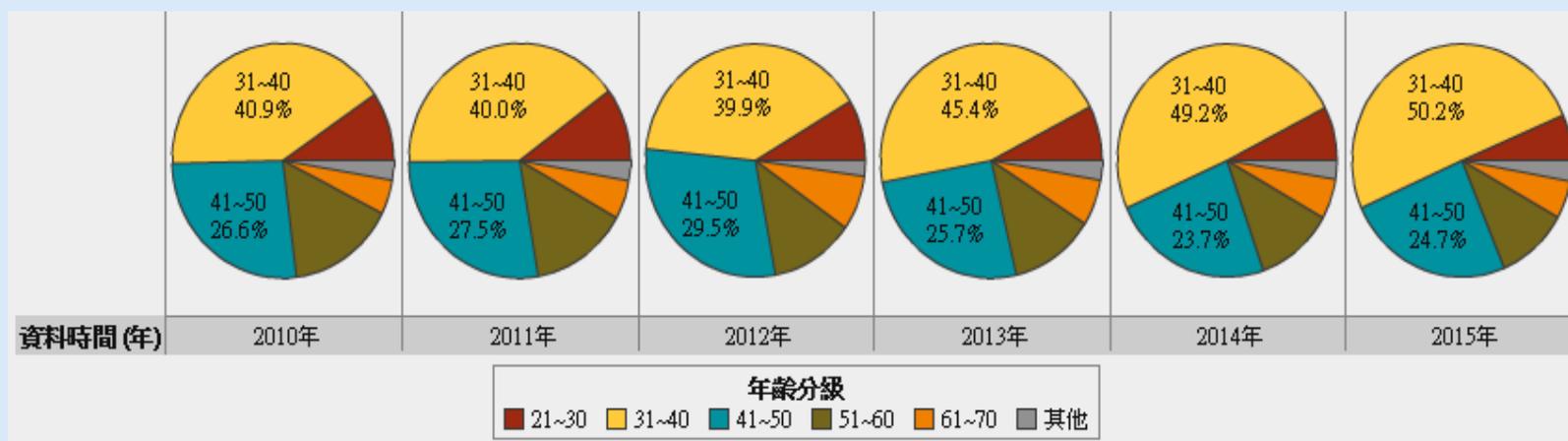
以漲停買進、跌停賣出占個股委託比重僅約1%，對股價波動影響程度甚微

成果分享(續)

權證交易 ①

各商品投資人平均年齡

年度	投資人平均年齡		
	股票	權證	ETF
2010年	46.43	40.50	43.65
2011年	46.76	40.43	44.37
2012年	47.32	40.83	44.63
2013年	48.13	41.49	45.24
2014年	48.29	41.54	46.18
2015年	48.48	41.46	46.78

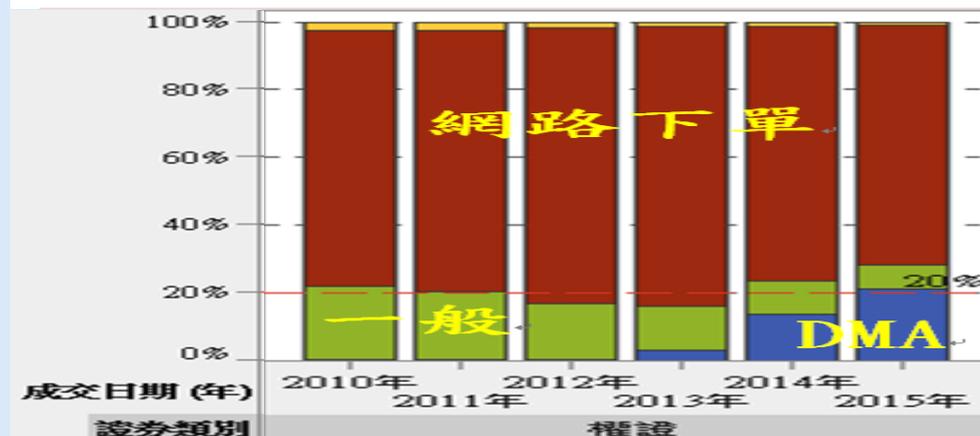


成果分享(續)

權證交易 ②

TOP 10 權證下單縣市(依成交金額)

下單券商位置	成交金額	分點數	投資人數
台北市	NT\$730,319,318,210.0	425	91,223
新北市	NT\$307,130,334,720.0	176	41,277
台中市	NT\$182,333,003,970.0	131	35,767
高雄市	NT\$155,456,445,850.0	152	30,056
桃園市	NT\$117,449,092,250.0	95	22,862
台南市	NT\$113,026,113,180.0	93	19,723
新竹市	NT\$77,241,195,080.0	43	14,184
彰化縣	NT\$44,527,791,490.0	48	9,743
嘉義市	NT\$30,199,109,640.0	29	4,665
新竹縣	NT\$22,176,187,210.0	18	4,351



成果分享(續)

權證交易 ③

大、中、小型發行人造市品質比較

- -大型
- -中型
- -小型

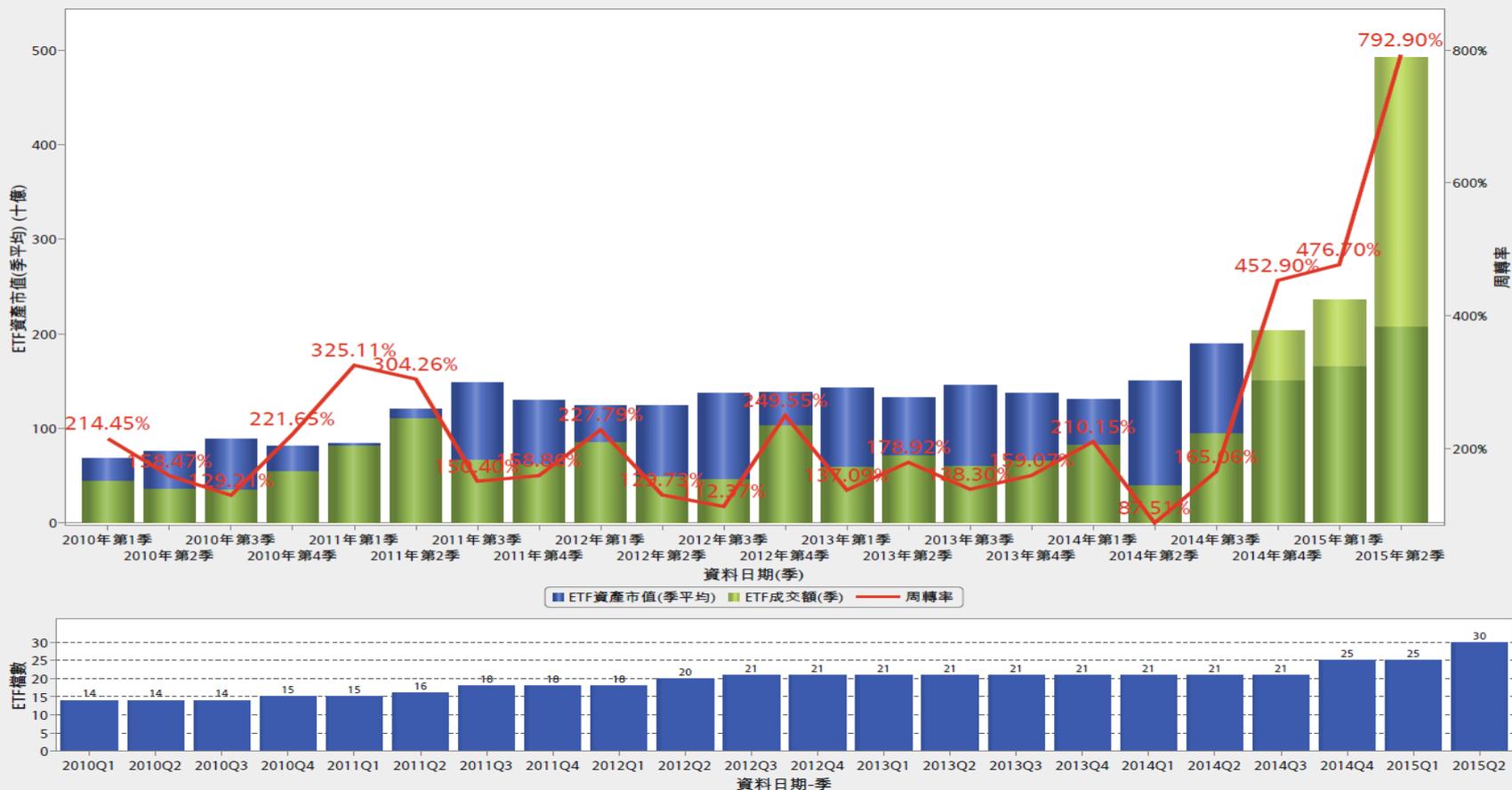
圓圈大小：委
託買進金額



ETF交易 ①

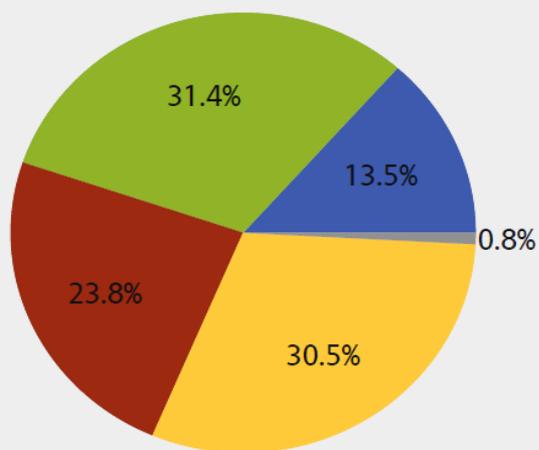


ETF市場概況



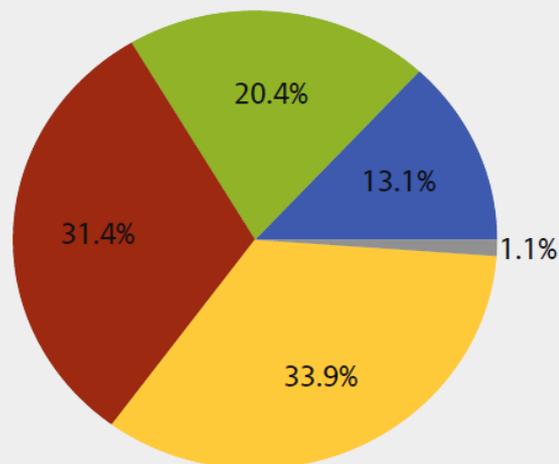
ETF交易 ②

2010年至2011年第二季
ETF成交金額



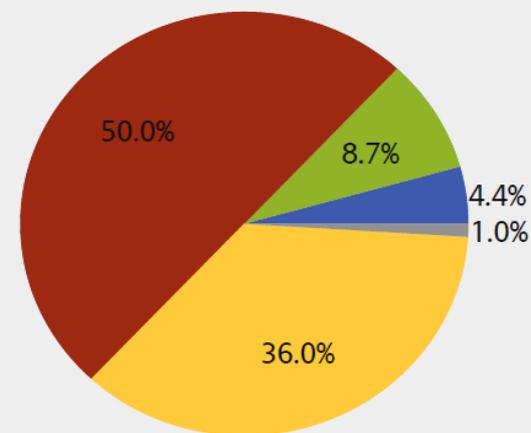
投資人類別
 ■ 外資 ■ 本國法人 ■ 自然人
 ■ 自營商 ■ 其他

2011年第三季至2014年第三季
ETF成交金額



投資人類別
 ■ 外資 ■ 本國法人 ■ 自然人
 ■ 自營商 ■ 其他

2014年第四季至2015年第二季
ETF成交金額

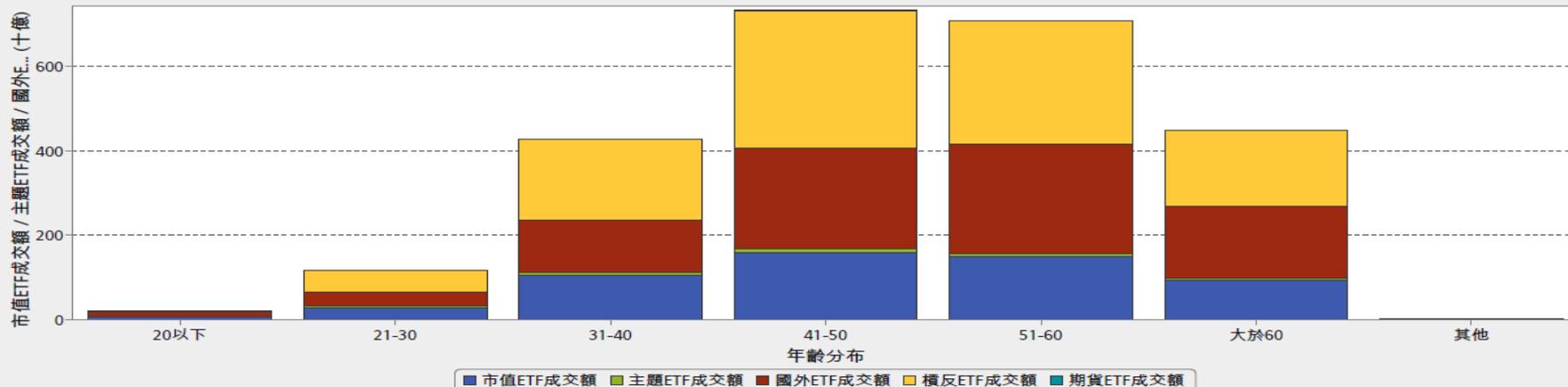


投資人類別
 ■ 外資 ■ 本國法人 ■ 自然人
 ■ 自營商 ■ 其他

ETF交易 ③

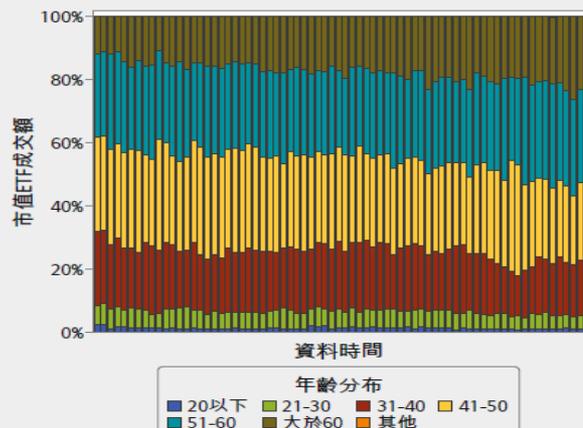
自然人年紀分布

各類型ETF之年紀分布



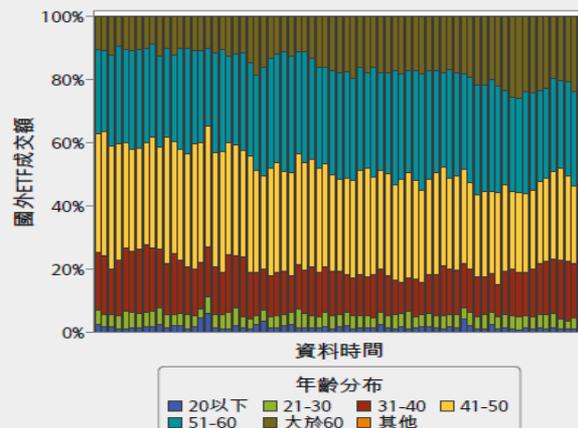
▼ A4.1

市值ETF之年紀分布



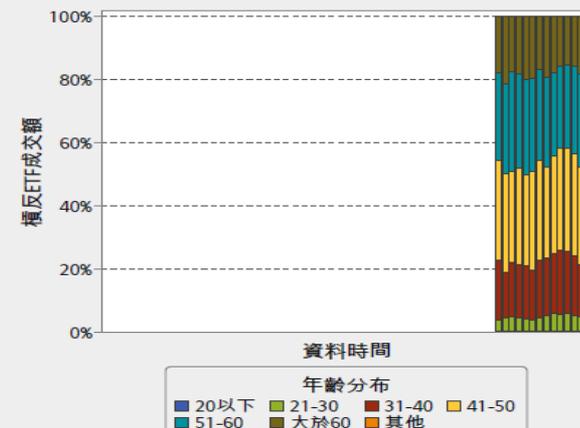
▼ A4.2

國外ETF之年紀分布



▼ A4.3

槓反ETF之年紀分布



▼ A4.4

結論

總經理主導

- 跨部門協調
- 進度追蹤控管
- 重要事項裁示



本案成功關鍵

①

電腦部門相互合作

- 原始資料彙整轉換
- 建置資料倉儲
- 業務協調及平台管理



結論(續)

業務部門自主分析

- 變更觀念
- 自主分析，調整結果



本案成功關鍵 ②

顧問專業協助

- 提供導入經驗及程序
- 協助過程進行
- 問題分析及解決



結論(續)

內部競賽

競賽
辦法

資料
彙整

教育
訓練

獎項

- 個人獎
- 部門獎

競賽方式

- 初賽
- 複賽

結論(續)

2016年大數據分析題目

證券市場異常資訊偵測系統之規劃與應用

- ①
 - 媒體、網路、社群等異常資訊偵測
 - 與個股價量波動、投資人交易資訊之整合



整合結構化與非結構化大數據，構建對上市公司管理預警系統之實證研究

- ②
 - 整合財報等結構化資料及產業資訊、媒體等非結構化資料
 - 構建對上市公司平時與例外管理預警指標及評鑑系統

結論(續)

2016年大數據分析題目

證券商財務數據分析與監理應用

- 蒐集長期財務數據及證券商經營情形變化，配合市場資訊進行分析，作為監理證券商之輔助資訊，或政策開放業務之參考
- ③
- 協助券商瞭解市場定位，提供專業分析



報告完畢 敬請指教

